# Optimal Observables for Minimum-Error State Discrimination in General Probabilistic Theories

Koji Nuida, Gen Kimura, Takayuki Miyadera

Research Center for Information Security (RCIS), National Institute of Advanced Industrial Science and Technology (AIST)
Akihabara-Daibiru Room 1003, 1-18-13 Sotokanda, Chiyoda-ku, Tokyo 101-0021, Japan
E-mail: {k.nuida, gen-kimura, miyadera-takayuki}[at]aist.go.jp

**Abstract**

General Probabilistic Theories provide the most general mathematical framework for the theory of probability in an operationally natural manner, and generalize classical and quantum theories. In this article, we study state-discrimination problems in general probabilistic theories using a Bayesian strategy. After re-formulation of the theories with mathematical rigor, we first prove that an optimal observable to discriminate any (finite) number of states always exists in the most general setting. Next, we revisit our recently proposed geometric approach for the problem and show that, for two-state discrimination, this approach is indeed effective in arbitrary dimensional cases. Moreover, our method reveals an operational meaning of Gudder's "intrinsic metric" by means of the optimal success probability, which turns out to be a generalization of the trace distance for quantum systems. As its by-product, an information-disturbance theorem in general probabilistic theories is derived, generalizing its well known quantum version.

## 1 Introduction

### 1.1 Background

Among many attempts to understand quantum theory axiomatically, an operationally natural approach for the general theory of probability, recently referred to as *general probabilistic theories* (or *generic probabilistic models*), has been studied [8, 9, 11, 16] and has attracted much attention in the recent development of quantum information theory (e.g., [1, 5, 13]). Such an approach provides a unified mathematical framework that involves not only classical and quantum theories but also more general settings that would be candidates of possible future extensions of the present quantum theory. One of the motivations of such an approach is to understand quantum mechanics better by introducing various viewpoints especially with information theoretic point of view. Another motivation to investigate such a general theory has arisen recently from research on quantum information theory including quantum information security. Among recent development of information theory and information security, one of the greatest impacts was provided by Shor's discovery [19] of an efficient (i.e., polynomial-time) integer factoring algorithm for quantum computers that reveals a future practical threat against several standard cryptosystems in the present time, such as RSA cryptosystem [17]. This history suggests a non-negligible possibility that any cryptosystem with security based on the present physical theory, even quantum theory, may fall into insecure once a further advanced physical theory is discovered and applied to information technology. Hence a study of possible extensions of the present physical theory is of importance and interest from not only theoretical but also practical standpoints.

One of the most important aims of studying general probabilistic theories is to determine which characteristics are typical for classical or quantum systems and which are not. For example, in a recent article [1] Barnum et al. investigated cloning and broadcasting of states in a general probabilistic theory. They proved (in finite-dimensional cases) that universal cloning or universal broadcasting is possible only for classical systems, which generalizes the No-Cloning Theorem and the No-Broadcasting Theorem for quantum systems [2, 6, 20, 21]. Another example relevant to our present work is our recent study [13] on minimum-error state discrimination problems in general probabilistic theories (in this article the word "minimum-error" is omitted since we do not discuss other kinds of discrimination problems such as unambiguous state discrimination). State discrimination problems have been well investigated for quantum systems (e.g., [3, 10, 12, 22]), but optimal success probabilities to discriminate given states and the corresponding optimal measurements were determined only in very restricted cases such as two-state cases. In [13] we gave a formulation of state discrimination problems in finite-dimensional general probabilistic theories, and introduced from a geometric viewpoint a class of special ensembles of states called *Helstrom families*: We showed that the optimal success probability can be determined by a Helstrom family if it exists. For the existence, we have discussed only for two-state cases and some other cases of states with symmetric configuration, and it has been shown that a Helstrom family always exists for both classical and quantum systems in any "generic" case (specified in a certain well-defined manner). However, existence of Helstrom families in more general (neither classical nor quantum) cases has not been clarified. The main aim of this article is to study the existence problem of the Helstrom family in general probabilistic theories with arbitrary dimension that are neither classical nor quantum.

### 1.2 Our contributions and organization of the article

In Sect. 2, we summarize a mathematical framework for general probabilistic theories. Following several preceding works for general probabilistic theories (e.g., [1, 9, 11, 13, 14, 16]), our formulation is based on the notions of *states*, *effects* and *observables*, as well as the notion of probabilistic state ensembles. Namely, we regard the state space as a "convex structure" [9]. A standard argument shows that the associated "separated" state space is embedded as a convex subset $\mathcal{S}$ in a real vector space $V$. For the sake of minimality, we assume that $V$ is the affine hull of $\mathcal{S}$ and the topology of $V$ is the weak topology generated by all effects on $\mathcal{S}$. We emphasize that $\mathcal{S}$ is usually assumed to be compact with respect to this topology, but in the present article compactness is not assumed to keep the most generality of our setting. In fact, when $\mathcal{S}$ is not compact with respect to this topology, we further take a "virtual state space" $\widetilde{\mathcal{S}} \supset \mathcal{S}$ and a "virtual underlying space" $\widetilde{V} \supset V$ such that $\widetilde{\mathcal{S}}$ is a *compact* convex subset of $\widetilde{V}$ and some additional conditions are satisfied (see Theorem 2.1 for the precise statement):

$$\begin{array}{ccccc} & \mathrm{cl}_{\widetilde{V}}(\mathcal{S}) = & \widetilde{\mathcal{S}} & \subset & \widetilde{V} \\ & & \cup & & \cup \\ \overline{\mathcal{S}_0} = \mathcal{S}_0/{\sim} & \simeq & \mathcal{S} & \subset & V \end{array}$$

By those properties, the objects $V$, $\widetilde{\mathcal{S}}$ and $\widetilde{V}$ are uniquely determined by $\mathcal{S}$, called the *minimal framework*. See Appendices A–E for further technical details. Now the effects on the "real" state space $\mathcal{S}$ are in one-to-one correspondence to their continuous extensions to $\widetilde{\mathcal{S}}$, called "virtual effects". A similar correspondence exists between observables on $\mathcal{S}$ and "virtual observables" on $\widetilde{\mathcal{S}}$. Moreover, for each "virtual state" $\widetilde{s} \in \widetilde{\mathcal{S}} \setminus \mathcal{S}$, any $\varepsilon > 0$ and any observables $\mathbf{O}_1, \ldots, \mathbf{O}_k$, there exists a "real state" $s \in \mathcal{S}$ such that the results of measurements of these $\mathbf{O}_i$ at $\widetilde{s}$ are within $\varepsilon$-error from the results at $s$; physically, this means that virtual states and real states are indistinguishable by experiments. Note that, in finite-dimensional cases, the underlying space $V$ is always isomorphic to a finite-dimensional Euclidean space and now $\mathcal{S}$ is nothing but a bounded convex subset of the Euclidean space $V$.

In Sect. 3, we give a natural formulation of state discrimination problems in general probabilistic theories by following our preceding work [13]. Our present formulation coincides with the preceding one when $\mathcal{S}$ is compact. Moreover, we show that an optimal observable always exists for discrimination of any (finite) number of given states with arbitrary a priori occurrence probabilities (see Theorem 3.1). Although it would be possible to interpret this result as a special case of a general theorem by Ozawa [16], we include the proof in this article for the reader's convenience because of its simplicity. (The proof uses only the existence theorem of maximum values of continuous functions on compact spaces and some elementary arguments for topological spaces.) Note that the argument in Sect. 3 is closed within the real state space $\mathcal{S}$, therefore the additional notions such as virtual states and virtual observables are not yet needed.

In Sect. 4, we introduce the notion of (weak) Helstrom families by translating the definition given in [13] to our minimal framework. A weak Helstrom family yields an upper bound of the optimal success probability for discriminating given states, while a Helstrom family yields the tight bound. A sufficient condition for a weak Helstrom family to be a Helstrom family has been given [13]. As a consequence of the above-mentioned existence theorem of an optimal observable, we show that the above sufficient condition is also necessary, except for the meaningless cases called *non-generic cases*. (By definition, *generic cases* are the cases where there exists a discrimination strategy better than simply outputting the candidate state with highest a priori probability.) In two-state cases, the above necessary and sufficient condition turns out to be "distinguishability" of two (possibly virtual) states $t_1, t_2$ associated to a given weak Helstrom family, therefore the problem of finding a Helstrom family is reduced to a study of distinguishable (virtual) states.

Finally, in Sect. 5 we prove that a Helstrom family for two-state discrimination always exists in generic cases (see Theorem 5.3), hence in such a case the optimal success probability can be determined (at least in principle) by just finding a Helstrom family. Our argument works in a general case of arbitrary dimension that may be neither classical nor quantum. Owing to the result, we also give a simple criterion for generic cases among all two-state cases (see Theorem 5.4): Given two distinct candidate states $s_1, s_2 \in \mathcal{S}$ with positive a priori probabilities $p_1, p_2$, the case is non-generic if and only if we have $p_1 \neq p_2$ and an element $s^* = (p_1 s_1 - p_2 s_2)/(p_1 - p_2)$ of $V$ lies outside the state space $\mathcal{S}$. In particular, the equiprobable cases $p_1 = p_2 = 1/2$ are always generic, therefore in such cases we are always able to discriminate (at least in principle) given states with probability higher than $1/2$. Moreover, our result also reveals a relation of Gudder's distance between two states $s_1, s_2 \in \mathcal{S}$ [9] with the optimal success probability of discriminating $s_1$ and $s_2$ in equiprobable cases, and also an operational meaning of Gudder's intrinsic metric [9] that gives an operationally natural generalization of the trace distance for quantum systems to general probabilistic theories (see Remark 5.1). As an application, a simple (qualitative) version of the information disturbance theorem in general probabilistic theories is shown to be hold that generalizes the corresponding theorem in quantum theory.

## 2 A Mathematical Framework for General Probabilistic Theories

In this section, we introduce a mathematical framework for general probabilistic theories. In this article, any vector space is defined over the real field $\mathbb{R}$ unless otherwise specified.

Following the preceding works [1, 9, 11, 13, 14, 16], we start with a set $\mathcal{S}_0$ of *states*, called a *state space*, that is a *convex structure* [9] in the following sense: For two states $s, t \in \mathcal{S}_0$ and two weights $\lambda, \mu \geq 0$ such that $\lambda + \mu = 1$, a state $\langle \lambda, \mu; s, t \rangle \in \mathcal{S}_0$ called an *ensemble* of $s, t$ with weights $\lambda, \mu$ is uniquely determined. Physically, $\langle \lambda, \mu; s, t \rangle$ means the probabilistic state ensemble of $s$ and $t$ with a priori probabilities $\lambda$ and $\mu$. We regard any convex subset of a vector space as a convex structure with a natural operation $\langle \lambda, \mu; s, t \rangle = \lambda s + \mu t$. Note that any other postulate for the operation $\langle \lambda, \mu; s, t \rangle$ is not required; some natural properties of state ensembles will be induced by construction of the associated "separated" state space presented below.

For any convex structure $C$, we say that a functional $f : C \to \mathbb{R}$ on $C$ is *affine* if we have $f(\langle \lambda, \mu; s, t \rangle) = \lambda f(s) + \mu f(t)$ for any $s, t \in C$. Let $\mathcal{E}(C)$ denote the set of all affine functionals $e$ on $C$ with image $e(C)$ contained in the unit interval $[0, 1]$ in $\mathbb{R}$. Then we call each $e \in \mathcal{E}(\mathcal{S}_0)$ an *effect* on $\mathcal{S}_0$. Now we define an equivalence relation $\sim$ on $\mathcal{S}_0$ by setting $s \sim t$ if and only if $e(s) = e(t)$ for every $e \in \mathcal{E}(\mathcal{S}_0)$. Let $\overline{s}$ denote the equivalence class of $s \in \mathcal{S}_0$. Then the quotient set $\overline{\mathcal{S}_0} = \mathcal{S}_0/\sim$ is also a convex structure with $\langle \lambda, \mu; \overline{s}, \overline{t} \rangle = \overline{\langle \lambda, \mu; s, t \rangle}$ for any $\overline{s}, \overline{t} \in \overline{\mathcal{S}_0}$. A physical interpretation is that, as two equivalent states (in the above sense) are statistically indistinguishable for any effect, we would have no physical way to distinguish those states. (See below for the definition of observables composed of effects.) Now each $e \in \mathcal{E}(\mathcal{S}_0)$ induces an effect $\overline{e} \in \mathcal{E}(\overline{\mathcal{S}_0})$ on $\overline{\mathcal{S}_0}$ by $\overline{e}(\overline{s}) = e(s)$ for each $\overline{s} \in \overline{\mathcal{S}_0}$, and this defines a one-to-one correspondence between $\mathcal{E}(\mathcal{S}_0)$ and $\mathcal{E}(\overline{\mathcal{S}_0})$. Moreover, the definition of the set $\overline{\mathcal{S}_0}$ implies the following property (see e.g., [9, 11, 16]):

**Lemma 2.1.** *The convex structure $\overline{\mathcal{S}_0}$ is separated, in the sense that for any distinct $s, t \in \overline{\mathcal{S}_0}$, there exists an effect $e \in \mathcal{E}(\overline{\mathcal{S}_0})$ such that $e(s) \neq e(t)$.*

The next theorem presents our framework involving the separated state space $\overline{\mathcal{S}_0}$. To our framework we intend to introduce as few mathematical structures as possible subject to physically natural requirements; we call the resulting framework a *minimal framework*. Here we use the notion of topological vector spaces; we refer to the book [18] for theory of topological vector spaces together with some relevant topics in general topology. In what follows, we abbreviate "topological vector space" to "t.v.s.", and "locally convex" to "l.c.". For any t.v.s. $W$, let $\mathcal{L}_c(W)$ denote the set of all continuous linear functionals $W \to \mathbb{R}$. Moreover, let $\mathcal{T}(X)$ denote the topology on a set $X$ if it is clear from the context. Then the above-mentioned theorem on our minimal framework is the following:

**Theorem 2.1.** *Given a separated convex structure $\overline{\mathcal{S}_0}$ as above, there exist the following objects:*

- *a l.c. Hausdorff t.v.s. $\widetilde{V}$ (over $\mathbb{R}$);*
- *a convex subset $\widetilde{\mathcal{S}}$ of $\widetilde{V}$ such that $\widetilde{V}$ is the affine hull $\mathrm{Aff}(\widetilde{\mathcal{S}})$ of $\widetilde{\mathcal{S}}$;*
- *a topological vector subspace $V$ of $\widetilde{V}$ that is dense in $\widetilde{V}$;*
- *a convex subset $\mathcal{S}$ of $V$ such that $\mathrm{Aff}(\mathcal{S}) = V$,*

*satisfying the following conditions:*

- *$\mathcal{S}$ is isomorphic to $\overline{\mathcal{S}_0}$, in the sense that there exists a bijection $\varphi : \overline{\mathcal{S}_0} \to \mathcal{S}$ such that $\varphi(\langle \lambda, \mu; s, t \rangle) = \lambda \varphi(s) + \mu \varphi(t)$ for any $s, t \in \overline{\mathcal{S}_0}$;*

- *the topology $\mathcal{T}(\widetilde{V})$ of $\widetilde{V}$ is a weak topology, i.e., the topology with minimal family of open subsets to make every $f \in \mathcal{L}_c(\widetilde{V})$ continuous;*
- *the induced topology on $\mathcal{S}$ is the weakest to make every $e \in \mathcal{E}(\mathcal{S})$ continuous;*
- *the induced topology on $V$ is the weakest to make every linear functional $f : V \to \mathbb{R}$, such that $f(\mathcal{S}) \subset \mathbb{R}$ is bounded, a continuous map;*
- *$\widetilde{\mathcal{S}}$ is the closure $\mathrm{cl}_{\widetilde{V}}(\mathcal{S})$ of $\mathcal{S}$ in $\widetilde{V}$, and $\widetilde{\mathcal{S}}$ is compact and complete.*

$$\begin{array}{rcccc} & \mathrm{cl}_{\widetilde{V}}(\mathcal{S}) = & \widetilde{\mathcal{S}} & \subset & \widetilde{V} \\ & & \cup & & \cup \\ \overline{\mathcal{S}_0} = \mathcal{S}_0/\!\sim & \simeq & \mathcal{S} & \subset & V \end{array}$$

*Moreover, these objects are unique; namely, for another collection $\mathcal{S}'$, $V'$, $\widetilde{\mathcal{S}}'$ and $\widetilde{V}'$ of such objects, there exists an affine isomorphism $\widetilde{V} \to \widetilde{V}'$ that is a homeomorphism and maps each of $\mathcal{S}$, $V$ and $\widetilde{\mathcal{S}}$ onto the corresponding object.*

A proof of Theorem 2.1 will be given in Appendices A–E.

*Remark* 2.1. In finite-dimensional cases ($\dim \mathcal{S} = n < \infty$), the space $V$ above is isomorphic to an $n$-dimensional Euclidean space $\mathbb{R}^n$ (cf., Theorem 5.1), and we have $\widetilde{V} = V$ and $\widetilde{\mathcal{S}} = \mathrm{cl}_V(\mathcal{S})$. Hence in such cases, the state space $\mathcal{S}$ is nothing but a bounded convex subset of $\mathbb{R}^n$. Moreover, in this case every $e \in \mathcal{E}(\widetilde{\mathcal{S}})$ is continuous by the definition of $\mathcal{T}(V) = \mathcal{T}(\widetilde{V})$; however, the continuity is not guaranteed in a general case.

**Definition 2.1.** We call the sets $\mathcal{S}$, $\widetilde{\mathcal{S}}$, $V$, and $\widetilde{V}$ a *(real) state space*, a *virtual state space*, a *(real) underlying space*, and a *virtual underlying space*, respectively. We call $s \in \mathcal{S}$ a *(real) state* and $\widetilde{s} \in \widetilde{\mathcal{S}} \setminus \mathcal{S}$ a *virtual state*. Moreover, we call each $e \in \mathcal{E} = \mathcal{E}(\mathcal{S})$ a *(real) effect* on $\mathcal{S}$, and each $\widetilde{e} \in \mathcal{E}(\widetilde{\mathcal{S}})$ a *virtual effect* on $\widetilde{\mathcal{S}}$ if it is continuous. Let $\widetilde{\mathcal{E}}$ denote the set of the virtual effects on $\widetilde{\mathcal{S}}$, i.e., $\widetilde{\mathcal{E}} = \{\widetilde{e} \in \mathcal{E}(\widetilde{\mathcal{S}}) \mid \widetilde{e} \text{ is continuous}\}$.

The choice of $\mathcal{T}(\mathcal{S})$ is motivated by a physical intuition that any available information on the state space $\mathcal{S}$ would be obtained via statistical properties of effects on $\mathcal{S}$. On the other hand, the continuity of virtual effects are required to ensure the following correspondence between effects and virtual effects:

**Lemma 2.2.** *Each effect $e \in \mathcal{E}$ on $\mathcal{S}$ has a unique continuous affine extension $\widetilde{e} : \widetilde{\mathcal{S}} \to \mathbb{R}$, and we have $\widetilde{e} \in \widetilde{\mathcal{E}}$. This gives a bijection $e \mapsto \widetilde{e}$ from $\mathcal{E}$ to $\widetilde{\mathcal{E}}$.*

*Proof.* Only the nontrivial part is the existence of a continuous affine extension $\widetilde{e}$ of $e$ with $\widetilde{e} \in \widetilde{\mathcal{E}}$; the uniqueness then follows since $\mathcal{S}$ is dense in $\widetilde{\mathcal{S}}$. First, since $\mathrm{Aff}(\mathcal{S}) = V$, the effect $e$ extends to an affine functional $f : V \to \mathbb{R}$. Let $\alpha$ be the value of $f$ at the origin of $V$; therefore $f' = f - \alpha : V \to \mathbb{R}$ is linear. Note that $f'(\mathcal{S}) \subset [-\alpha, 1 - \alpha]$, therefore $f'$ is continuous on $V$ by the property of $V$ in Theorem 2.1. By a consequence of Hahn-Banach's Theorem (Theorem D.1), this $f'$ extends to a continuous linear functional $g$ on $\widetilde{V}$. Now $g(\widetilde{\mathcal{S}}) \subset \mathrm{cl}_{\mathbb{R}}(g(\mathcal{S}))$ since $\widetilde{\mathcal{S}} = \mathrm{cl}_{\widetilde{V}}(\mathcal{S})$, while $g(\mathcal{S}) \subset [-\alpha, 1 - \alpha]$ since $g$ is an extension of $f'$. Thus the restriction $\widetilde{e} = (g + \alpha)|_{\widetilde{\mathcal{S}}}$ of $g + \alpha$ to $\widetilde{\mathcal{S}}$ is a continuous affine functional such that $\widetilde{e}(\widetilde{\mathcal{S}}) \subset [0, 1]$, therefore $\widetilde{e} \in \widetilde{\mathcal{E}}$. This $\widetilde{e}$ is the desired extension of $e$. □

Moreover, the sets $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ have the following properties:

**Lemma 2.3.** *Both $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ are separated, which (for $\widetilde{\mathcal{S}}$) means that for any distinct $s, t \in \widetilde{\mathcal{S}}$, there exists an $e \in \widetilde{\mathcal{E}}$, not just $e \in \mathcal{E}(\widetilde{\mathcal{S}})$, such that $e(s) \neq e(t)$.*

*Proof.* Since $V$ is Hausdorff, $\mathcal{S}$ is separated (in the sense of Lemma 2.1) by the definition of $\mathcal{T}(\mathcal{S})$; see Theorem 2.1. On the other hand, let $s, t$ be distinct elements of $\widetilde{\mathcal{S}}$. Then, since $\mathcal{T}(\widetilde{V})$ is Hausdorff and a weak topology (see Theorem 2.1), there exists a continuous linear functional $f$ on $\widetilde{V}$ such that $f(s) \neq f(t)$. Now $f(\widetilde{\mathcal{S}}) \subset \mathbb{R}$ is bounded since $\widetilde{\mathcal{S}}$ is compact, therefore the restriction $e$ of an appropriate affine transformation $\alpha f + \beta$ of $f$ to $\widetilde{\mathcal{S}}$, where $\alpha, \beta \in \mathbb{R}$, is a virtual effect such that $e(s) \neq e(t)$. Hence Lemma 2.3 holds. □

**Definition 2.2.** An $N$-valued *(real) observable* (or *virtual observable*, respectively) is a collection $\mathbf{O} = (e_i)_{i=1}^{N}$ of $N$ effects $e_i \in \mathcal{E}$ (or $N$ virtual effects $e_i \in \widetilde{\mathcal{E}}$, respectively) such that $\sum_{i=1}^{N} e_i = 1$. Let $\mathcal{O}_N$ and $\widetilde{\mathcal{O}}_N$ denote the sets of all $N$-valued observables and of all $N$-valued virtual observables, respectively.

Physically, for each observable $\mathbf{O} = (e_i)_i$ and each $s \in \mathcal{S}$, the quantity $e_i(s)$ means the probability to obtain $i$-th output when measuring $\mathbf{O}$ at the state $s$; the condition $\sum_i e_i = 1$ is required by a property of probability. On the other hand, the affine property of each $e_i$ is motivated by a natural expectation that the output probabilities for a probabilistic state ensemble would be weighted sums of those probabilities for each of the original state. The same also holds for virtual observables. Now we have the following correspondence:

**Lemma 2.4.** *We have $\widetilde{\mathbf{O}} = (\widetilde{e_i})_i \in \widetilde{\mathcal{O}}_N$ for any $\mathbf{O} = (e_i)_i \in \mathcal{O}_N$. This gives a bijection $\mathbf{O} \mapsto \widetilde{\mathbf{O}}$ from $\mathcal{O}_N$ to $\widetilde{\mathcal{O}}_N$.*

*Proof.* Only the nontrivial part is to show that $\sum_i \widetilde{e_i} = 1$ for any $\mathbf{O} = (e_i)_i \in \mathcal{O}_N$. This follows from the uniqueness property in Lemma 2.2, since both $\sum_i \widetilde{e_i}$ and 1 are continuous affine extensions of the effect $\sum_i e_i = 1$ to $\widetilde{\mathcal{S}}$. □

By virtue of Lemma 2.4, the output probabilities for virtual observables at virtual states can be derived (at least in principle) from information on real observables at real states. On the other hand, for any finite collection of measurements with non-ideal accuracy, virtual states are indistinguishable from real states (in the sense mentioned in Sect. 1.2).

Note that our framework presented above does in fact not concern every feature of quantum theory, e.g., transformations of states possibly caused by measuring observables. However, our framework is still enough for our current purpose of studying state discrimination problems.

Obviously, two fundamental examples of general probabilistic theories are given by classical and quantum theories, as follows (taken from [13]):

*Example* 2.1. A finite classical system described by a finite probability theory with finite sample space $\{\omega_1, \dots, \omega_n\}$ is formulated in our model as the $(n-1)$-dimensional standard simplex $\mathcal{S} = \{p = (p_1, \dots, p_n) \in \mathbb{R}^n \mid p_i \geq 0, \sum_i p_i = 1\}$. Namely, each state is a probability distribution over the sample space, and it can be seen as a probabilistic ensemble of "pure states" $p^{(i)}$, $i = 1, \dots, n$, with only one possible output $\omega_i$, that are extremal points of $\mathcal{S}$ in usual sense. Note that in this example $\mathcal{S}$ itself is compact, hence all states are real. This example can be naturally extended to infinite-dimensional classical systems.

*Example* 2.2. In quantum theory, a quantum state is described by a density operator $\rho$, that is a positive operator on a Hilbert space $\mathcal{H}$ with unit trace. Thus the state space is a convex subset of the vector space of all linear operators on $\mathcal{H}$. Moreover, an effect $e$ is described [16] by a positive bounded operator $B$ such that $0 \leq B \leq I_{\mathcal{H}}$ via the relation $e(\rho) = \mathrm{tr} B\rho$, that is an element of positive operator valued measure (POVM).

In the last of this section, we give two remarks on relations with preceding works. Before starting the remarks, note that assumptions on compactness of the state space $\mathcal{S}$ and on completeness of $\mathcal{S}$ are equivalent to each other, since each of the two implies that $\mathcal{S}$ is closed in $\widetilde{V}$ and hence $\widetilde{\mathcal{S}} = \mathcal{S}$.

*Remark* 2.2. In a recent work by Barnum et al. [1], their finite-dimensional state space is assumed to be compact to guarantee that the state space is the closed convex hull of the set of "pure states" (i.e., extremal points of the state space). Owing to Krein-Milman's Theorem (see e.g., Theorem 10.4 in [18, Chapter II]), the same property is possessed by our (possibly infinite-dimensional) virtual state space $\widetilde{\mathcal{S}}$. Thus it is very attractive to start our argument by choosing the compact set $\widetilde{\mathcal{S}}$ as a new "state space" instead of $\mathcal{S}$. However, such a modification *does* decrease the generality of our framework. Namely, it is *not* guaranteed in general that every $e \in \mathcal{E}(\widetilde{\mathcal{S}})$, that should be a new "effect" in the above modification, is continuous with respect to the original topology of $\widetilde{\mathcal{S}}$. Thus to ensure that every "effect" is continuous, we need a new topology stronger than the original, therefore the set $\widetilde{\mathcal{S}}$ may fail compactness with respect to the new topology. Hence the advantage to choose $\widetilde{\mathcal{S}}$ as a state space disappears.

*Remark* 2.3. In another previous work by Gudder [9], the following distance $d(s,s')$ of two states $s,s' \in \mathcal{S}$ was introduced to make $\mathcal{S}$ a metric space. Namely, Gudder defined $d(s,s')$ to be the infimum of the values $0 < \lambda \leq 1$ such that $\lambda t + (1-\lambda)s = \lambda t' + (1-\lambda)s'$ for some states $t,t' \in \mathcal{S}$. Since this relation implies that $(1-\lambda)|e(s) - e(s')| = \lambda|e(t) - e(t')| \leq \lambda$ for any $e \in \mathcal{E}$, every effect is continuous with respect to the metric $d$ on $\mathcal{S}$. However, unless $\mathcal{S}$ is finite-dimensional, the metric $d$ is *not* necessarily continuous with respect to the topology of $\mathcal{S}$ specified in Theorem 2.1. This is roughly because, for a state $s \in \mathcal{S}$ and any collection of a finite number of effects $e_i$, the metric $d$ is not necessarily bounded by a sufficiently small value on the intersection of hyperplanes containing $s$ defined by the affine functionals $e_i$. Thus our topology on $\mathcal{S}$ is weaker than (or equal to) the topology defined by the metric $d$. Moreover, another relation of our results with Gudder's metric functions will be mentioned later (Remark 5.1).

# 3 State Discrimination Problems

In this section, we give a formulation of (minimum-error) state discrimination problems in general probabilistic theories based on the minimal framework introduced in Sect. 2. This formulation is a natural generalization of state discrimination problems for quantum systems, and in fact a naive translation of our preceding formulation [13] to the present more general setting.

In the state discrimination problem, we are given a finite number (say $N$) of real states $s_1, \dots, s_N \in \mathcal{S}$ and the corresponding a priori probabilities $p_1, \dots, p_N$, $p_i \geq 0$, $\sum_i p_i = 1$. To avoid inessential intricacy, we assume that each probability $p_i$ is positive. Then for each $N$-valued observable $\mathbf{O} = (e_i)_i \in \mathcal{O}_N$, we define the *success probability* $P_{\mathrm{succ}}(\mathbf{O})$ for the observable $\mathbf{O}$ by

$$P_{\mathrm{succ}}(\mathbf{O}) = \sum_{i=1}^{N} p_i e_i(s_i) \quad . \tag{1}$$

Namely, when measuring the observable $\mathbf{O}$ at an unknown state that is chosen from $s_1, \dots, s_N$ with probabilities $p_1, \dots, p_N$ (thus the unknown state is regarded as the probabilistic ensemble $\sum_i p_i s_i \in \mathcal{S}$), $i$-th output for $\mathbf{O}$ corresponds to the guess that the chosen state was originally $s_i$. (Without loss of generality, it suffices to consider $N$-valued observables when discriminating $N$ states.) Our aim is to make the success probability as high as possible. The *optimal success probability* $P_{\mathrm{succ}}$ is obviously defined by

$$P_{\mathrm{succ}} = \sup_{\mathbf{O} \in \mathcal{O}_N} P_{\mathrm{succ}}(\mathbf{O}) \quad , \tag{2}$$

and an observable $\mathbf{O} \in \mathcal{O}_N$ is called *optimal* if it attains the supremum, namely: $P_{\mathrm{succ}}(\mathbf{O}) = P_{\mathrm{succ}}$. However, it is nontrivial whether or not an optimal observable exists in each case. Ozawa

[16] has proven existence of Bayes optimal measurements under somewhat different formulation. The existence theorem also holds in our situation. Here we present the theorem together with its proof that is significantly simpler than the one in [16], as follows:

**Theorem 3.1.** *The supremum in the right-hand side of* (2) *is attained by an observable* $\mathbf{O} \in \mathcal{O}_N$*. Hence an optimal observable always exists.*

The rest of this section is devoted to the proof of Theorem 3.1; note that in this proof, virtual states do not appear at all. The outline is the following: With respect to a certain topology, the set $\mathcal{O}_N$ of $N$-valued observables is compact and the map $\mathcal{O}_N \to \mathbb{R}$, $\mathbf{O} \mapsto P_{\mathrm{succ}}(\mathbf{O})$, is continuous, therefore this map takes the maximum value at some $\mathbf{O} \in \mathcal{O}_N$. Now we introduce a map $\iota : \mathcal{E} \to [0,1]^{\mathcal{S}}$ from $\mathcal{E}$ to the direct product $[0,1]^{\mathcal{S}} = \prod_{s \in \mathcal{S}} [0,1]_s$ of copies $[0,1]_s$ of the unit interval $[0,1]$ over all $s \in \mathcal{S}$ by $\iota(e) = (e(s))_{s \in \mathcal{S}}$ for any $e \in \mathcal{E}$. Then $\iota$ is injective, therefore $\mathcal{E}$ is identified with the topological subspace $\iota(\mathcal{E})$ of the product space $[0,1]^{\mathcal{S}}$. By the definition of product topology, $\mathcal{T}([0,1]^{\mathcal{S}})$ is the weakest topology to make every projection $\pi_s : [0,1]^{\mathcal{S}} \to [0,1]_s$ $(s \in \mathcal{S})$ continuous. Thus the topology on $\mathcal{E}$ induced by the identification is the weakest to make every "evaluation map" $\mathsf{ev}_s : \mathcal{E} \to [0,1]$, $\mathsf{ev}_s(e) = e(s)$ $(s \in \mathcal{S})$ continuous. Now the following holds:

**Lemma 3.1.** $\iota(\mathcal{E})$ *is a closed subset of* $[0,1]^{\mathcal{S}}$*.*

*Proof.* For each $s, t \in \mathcal{S}$ and $0 \leq \lambda \leq 1$, put $s' = \lambda s + (1-\lambda)t \in \mathcal{S}$, and let

$$\mathcal{A}_{s,t,\lambda} = \{ f \in [0,1]^{\mathcal{S}} \mid \pi_{s'}(f) - \lambda \pi_s(f) - (1-\lambda)\pi_t(f) = 0 \} \ .$$

Then $\mathcal{A}_{s,t,\lambda}$ is a closed subset of $[0,1]^{\mathcal{S}}$, since the function $\pi_{s'} - \lambda \pi_s - (1-\lambda)\pi_t$ on $[0,1]^{\mathcal{S}}$ is continuous. Moreover, the affine property of the effects implies that $\iota(\mathcal{E})$ is the intersection of all the subsets $\mathcal{A}_{s,t,\lambda}$. Hence $\iota(\mathcal{E})$ is also closed in $[0,1]^{\mathcal{S}}$, therefore Lemma 3.1 holds. □

By Tychonoff's Theorem, the product space $[0,1]^{\mathcal{S}}$ is compact, therefore $\mathcal{E}$ is also compact with respect to the above topology by Lemma 3.1. Thus the product space $\mathcal{E}^N$ is also compact owing to Tychonoff's Theorem again. Moreover, a similar argument implies that the subset $\mathcal{O}_N$ of $\mathcal{E}^N$ is closed in $\mathcal{E}^N$, since the map $\mathcal{E}^N \to \mathbb{R}$, $(e_i)_{i=1}^N \mapsto \sum_{i=1}^N e_i(s)$, is continuous for every $s \in \mathcal{S}$. Thus $\mathcal{O}_N$ is also compact. Finally, with respect to the topology on $\mathcal{O}_N$, the above function $\mathbf{O} \mapsto P_{\mathrm{succ}}(\mathbf{O})$ on $\mathcal{O}_N$ is continuous. Hence the proof of Theorem 3.1 is concluded.

# 4 Helstrom Families

In Sect. 3, we have seen that an optimal observable to discriminate given states always exists in general probabilistic theories. In the quantum cases, the state discrimination problem has been intently investigated (e.g., [3, 10, 12, 22]), but strategies for attaining optimal solutions have been well established only in restricted cases such as two-state cases (cf., [10]) and some symmetric cases (cf., [3]). To study this problem in general probabilistic theories, our preceding work [13] introduced and studied the notion of "(weak) Helstrom families" from a geometric viewpoint. In this section, we give a translation of the preceding formulation to our minimal framework.

Recall that we are given $N$ states $s_i \in \mathcal{S}$ with a priori probabilities $p_i > 0$, $\sum_i p_i = 1$. Then the definition of weak Helstrom families is the following (cf., Definition 1 in [13]):

**Definition 4.1.** We call a family of $N$ ensembles $(\widetilde{p}_i, s_i; 1-\widetilde{p}_i, t_i)$, $i = 1, \dots, N$, a *weak Helstrom family*, if there exist a quantity $p \geq \max_i p_i$ called a *Helstrom ratio*, $N$ real or virtual states

$t_i \in \widetilde{\mathcal{S}}$, $i = 1, \ldots, N$ called *conjugate states* to $s_i$, and a real or virtual state $s \in \widetilde{\mathcal{S}}$ called a *reference state*, such that

$$\widetilde{p}_i s_i + (1 - \widetilde{p}_i) t_i = s \ , \quad \text{with } 0 < \widetilde{p}_i = \frac{p_i}{p} \leq 1 \tag{3}$$

for every $i$. We call a weak Helstrom family *trivial* when $p \geq 1$, and *nontrivial* when $p < 1$.

*Example* 4.1. In Fig. 1, we consider the case $N = 3$ and $p_i = 1/3$ $(i = 1, 2, 3)$. The three states $t_1, t_2, t_3$ are in such positions that their configuration is similar to that of $s_1, s_2, s_3$ with respect to the center $s$ of similarity, with similarity ratio $\overline{t_i s}/\overline{s_i s} = 2/1$. Now these form a weak Helstrom family with $\widetilde{p}_i = 2/3$, therefore the Helstrom ratio is $p = p_i/\widetilde{p}_i = 1/2$. Note that any other similar configuration with a larger similarity ratio gives a weak Helstrom family with larger $\widetilde{p}_i$, hence with a smaller Helstrom ratio.

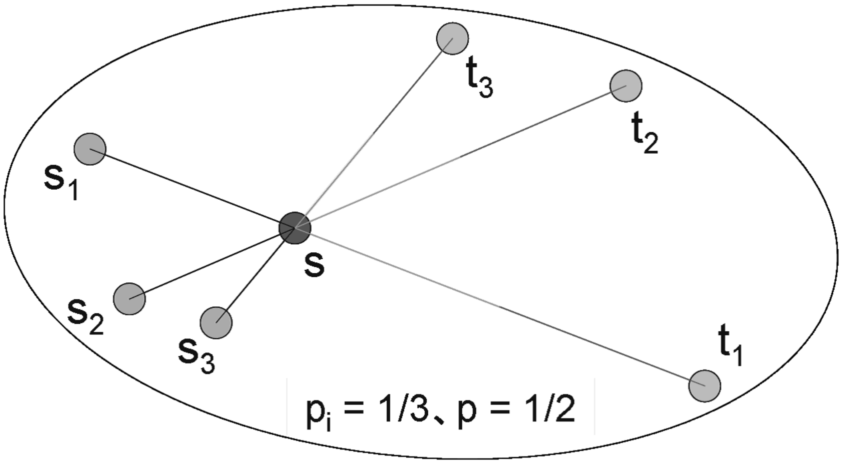

Figure 1: Example of weak Helstrom family

In the original paper [13], a weak Helstrom family was required to satisfy an additional condition $p \leq 1$, but here we relax this condition to simplify the argument. Note that a trivial weak Helstrom family with Helstrom ratio $p = 1$ always exists, by taking conjugate states $t_i = (1 - p_i)^{-1} \sum_{j \neq i} p_j s_j$ and a reference state $s = \sum_i p_i s_i$. Example 4.1 suggests that, intuitively, some nontrivial weak Helstrom families can be found as well by taking the states $t_i$ with larger configuration (cf., [13]). An importance of weak Helstrom families in a study of state discrimination problems is implied by the following property that has been proven in [13] under the framework there:

**Proposition 4.1** (cf., Proposition 1 in [13])**.** *For any weak Helstrom family with Helstrom ratio $p$, we have $P_{\text{succ}} \leq p$ for the optimal success probability.*

*Proof.* The idea of proof is essentially the same as [13]. For any observable $\mathbf{O} = (e_i)_i \in \mathcal{O}_N$

with the corresponding virtual observable $\widetilde{\mathbf{O}} = (\widetilde{e_i})_i \in \widetilde{\mathcal{O}}_N$ (see Lemma 2.4), we have

$$\begin{aligned} 1 &= \sum_{i=1}^{N} \widetilde{e_i}(s) \qquad (\text{using } \textstyle\sum_i \widetilde{e_i} = 1) \\ &= \sum_i \widetilde{e_i}(\widetilde{p_i} s_i + (1-\widetilde{p_i}) t_i) \qquad (\text{using (3)}) \\ &= \sum_i \widetilde{p_i} e_i(s_i) + \sum_i (1-\widetilde{p_i}) \widetilde{e_i}(t_i) \qquad (\text{using } s_i \in \mathcal{S}) \\ &= \sum_i \frac{p_i}{p} e_i(s_i) + \sum_i (1-\widetilde{p_i}) \widetilde{e_i}(t_i) \qquad (\text{using (3)}) \\ &= \frac{P_{\text{succ}}(\mathbf{O})}{p} + \sum_i (1-\widetilde{p_i}) \widetilde{e_i}(t_i) \qquad (\text{using (1)}). \end{aligned}$$

Since $\widetilde{p_i} \leq 1$, the second term of the last row is nonnegative, therefore we have $P_{\text{succ}}(\mathbf{O}) \leq p$ for any $\mathbf{O} \in \mathcal{O}_N$. Hence Proposition 4.1 holds. □

Note that the bound $P_{\text{succ}} \leq p$ given by Proposition 4.1 is meaningless when the weak Helstrom family is trivial. Thus only the weak Helstrom families that are significant for our purpose are the nontrivial ones. Now it was mentioned in Example 4.1 that changing the configuration of conjugate states to larger one makes the Helstrom ratio smaller, hence makes the bound given by Proposition 4.1 closer to the tight one. We are interested in whether or not the tight bound can be achieved just by this strategy. Owing to the observation, a notion of "Helstrom families", that is a special subclass consisting of "optimal" weak Helstrom families, was introduced in [13]:

**Definition 4.2** (cf., Definition 2 in [13])**.** We call a weak Helstrom family a *Helstrom family* if the Helstrom ratio $p$ attains the tight bound: $p = P_{\text{succ}}$.

If a Helstrom family exists, then we can determine (at least in principle) the optimal success probability by only searching (weak) Helstrom families by a certain (for example, geometric) method. However, existence of Helstrom families has been proven in the original work [13] only for some restricted cases. In this article, we investigate existence of Helstrom families in more general situations.

For this purpose, it is worthy to study conditions for a weak Helstrom family to be a Helstrom family. For one direction, a sufficient condition has been given in [13]. Here we prove the same result under the present framework:

**Proposition 4.2** (cf., Proposition 2 in [13])**.** *A sufficient condition for a weak Helstrom family* $(\widetilde{p_i}, s_i; 1-\widetilde{p_i}, t_i)$*,* $i = 1, \dots, N$*, to be a Helstrom family is that there exists* $\mathbf{O} = (e_i)_{i=1}^{N} \in \mathcal{O}_N$ *such that* $\widetilde{e_i}(t_i) = 0$ *for every* $i$*. Moreover, such an observable* $\mathbf{O}$ *is optimal (if exists):* $P_{\text{succ}}(\mathbf{O}) = P_{\text{succ}}$*.*

*Proof.* The idea is again the same as [13]. For such an observable $\mathbf{O}$, the argument in the proof of Proposition 4.1 implies that

$$1 = \frac{P_{\text{succ}}(\mathbf{O})}{p} + \sum_i (1-\widetilde{p_i}) \widetilde{e_i}(t_i) = \frac{P_{\text{succ}}(\mathbf{O})}{p} \ ,$$

hence $P_{\text{succ}}(\mathbf{O}) = p$. Now we have $P_{\text{succ}}(\mathbf{O}) \leq P_{\text{succ}} \leq p = P_{\text{succ}}(\mathbf{O})$ by Proposition 4.1, therefore $P_{\text{succ}}(\mathbf{O}) = P_{\text{succ}} = p$. Hence Proposition 4.2 holds. □

Again, Helstrom families are closely related to optimal state discrimination via Proposition 4.2. In the special case of two-state discrimination (i.e., $N = 2$), the above condition is rephrased as follows. Here we use the following terminology:

**Definition 4.3.** Two real or virtual states $t_1, t_2 \in \widetilde{\mathcal{S}}$ are said to be *distinguishable* if there exists an $\widetilde{e} \in \widetilde{\mathcal{E}}$ such that $\widetilde{e}(t_1) = 1$ and $\widetilde{e}(t_2) = 0$, i.e., the virtual observable $(1 - \widetilde{e}, \widetilde{e}) \in \widetilde{\mathcal{O}}_2$ discriminates $t_1$ and $t_2$ *with certainty.*

Then the rephrased condition is the following:

**Corollary 4.1** (cf., Theorem 1 in [13])**.** *Let $(\widetilde{p}_i, s_i; 1-\widetilde{p}_i, t_i)$, $i = 1, 2$, be a weak Helstrom family for two states $s_1, s_2 \in \mathcal{S}$ with a priori probabilities $p_1, p_2$. If the conjugate states $t_1$ and $t_2$ are distinguishable, then this weak Helstrom family is a Helstrom family. Moreover, an optimal observable $\mathbf{O}$ is given by an effect $e$ with the corresponding virtual effect $\widetilde{e}$ distinguishing $t_1$ and $t_2$: $\mathbf{O} = (1 - e, e)$.*

Now owing to the existence of an optimal observable (Theorem 3.1), we obtain a “converse” of the above facts. To state the result precisely, we recall the following notion introduced in [13]:

**Definition 4.4** ([13])**.** By *generic case* we signify any case in which the optimal success probability satisfies $P_{\text{succ}} > \max_i p_i$, and by *non-generic case* we signify any of the remaining cases, i.e., $P_{\text{succ}} = \max_i p_i$.

This definition means that, in non-generic cases, an optimal observable is always given by the trivial one that always returns $i$-th output with the index $i$ determined by $p_i = \max_j p_j$; namely, we always guess that a given state would be the most frequent $s_i$. Hence the state discrimination problem is nontrivial only in generic cases. Now we give the following result stating that the sufficient condition in Proposition 4.2 is also necessary in generic cases:

**Proposition 4.3.** *Let $(\widetilde{p}_i, s_i; 1 - \widetilde{p}_i, t_i)$, $i = 1, \ldots, N$, be a Helstrom family. Then, in generic cases, an optimal observable $\mathbf{O} = (e_i)_i \in \mathcal{O}_N$ for discriminating given states satisfies $\widetilde{e_i}(t_i) = 0$ for every $i$.*

*Proof.* For any optimal observable $\mathbf{O} = (e_i)_i$, since $P_{\text{succ}}(\mathbf{O}) = P_{\text{succ}} = p$, the argument in Proposition 4.1 implies that

$$1 = \frac{P_{\text{succ}}(\mathbf{O})}{p} + \sum_i (1 - \widetilde{p}_i)\widetilde{e_i}(t_i) = 1 + \sum_i (1 - \widetilde{p}_i)\widetilde{e_i}(t_i) \ ,$$

therefore $\sum_i (1 - \widetilde{p}_i)\widetilde{e_i}(t_i) = 0$. Thus we have either $\widetilde{p}_i = 1$ for some $i$, or $\widetilde{e_i}(t_i) = 0$ for every $i$. Now if $\widetilde{p}_i = 1$, then $P_{\text{succ}} = p = p_i/\widetilde{p}_i = p_i$, contradicting the assumption that we are in a generic case. Hence Proposition 4.3 holds. □

**Corollary 4.2.** *Let $(\widetilde{p}_i, s_i; 1 - \widetilde{p}_i, t_i)$, $i = 1, 2$, be a Helstrom family for two states $s_1, s_2$ with a priori probabilities $p_1, p_2$. Then, in generic cases, the conjugate states $t_1$ and $t_2$ are distinguishable by a virtual effect $\widetilde{e} \in \widetilde{\mathcal{E}}$ corresponding to an optimal observable $\mathbf{O} = (1 - e, e) \in \mathcal{O}_2$ for discriminating the states $s_1$ and $s_2$.*

*Proof.* By Proposition 4.3, an optimal observable $\mathbf{O} = (1-e, e) \in \mathcal{O}_2$ satisfies that $(1-\widetilde{e})(t_1) = 0$ and $\widetilde{e}(t_2) = 0$, therefore $\widetilde{e}(t_1) = 1$. □

# 5 Existence of Helstrom Families for Two-State Cases

In Sect. 4, we have presented some properties of (weak) Helstrom families for $N$-state cases. However, existence of Helstrom families has not been clarified so far. In this section, we investigate existence of Helstrom families particularly in two-state cases, i.e., $N = 2$. Note that our argument in this section works in a general setting, *not* necessarily classical or quantum, and also is *not* restricted to finite-dimensional cases.

Throughout this section, fix states $s_1, s_2 \in \mathcal{S}$ and a priori probabilities $p_1, p_2$. For simplicity, we assume that $s_1 \neq s_2$ and $p_1 \geq p_2$ by symmetry. Any (weak) Helstrom family in this section is for $s_1, s_2$ and $p_1, p_2$ unless otherwise specified.

## 5.1 A condition for generic cases

In this subsection, we present a condition for generic cases for later use. First we introduce an element $s^*$ of $V$ that plays a significant role in our following argument. Recall that we have assumed $p_1 \geq p_2$. If $p_1 > p_2$, then define

$$s^* = \frac{p_1 s_1 - p_2 s_2}{p_1 - p_2} = s_1 + \frac{p_1}{p_1 - p_2}(s_1 - s_2) \ .$$

Note that $s^* \in V$ since $s_1$ and $s_2$ are real states, therefore we have $s^* \in \widetilde{\mathcal{S}}$ if and only if $s^* \in \mathrm{cl}_V(\mathcal{S})$. Then the aforementioned condition is the following:

**Lemma 5.1.** *1. If the following condition*

$$\textit{either } p_1 = p_2, \textit{ or } p_1 > p_2 \textit{ and } s^* \notin \widetilde{\mathcal{S}} \tag{4}$$

*is satisfied and a Helstrom family exists, then it is a generic case.*

*2. If $p_1 > p_2$ and $s^* \in \widetilde{\mathcal{S}}$, then it is a non-generic case.*

*Proof.* Note that for any Helstrom family $(\widetilde{p}_i, s_i; 1-\widetilde{p}_i, t_i)$, $i = 1, 2$, it is a non-generic case if and only if $\widetilde{p}_1 = 1$ (since $p_1 \geq p_2$). Now if $p_1 = p_2$ and a Helstrom family exists, then $\widetilde{p}_1 = 1$ implies that $\widetilde{p}_2 = 1$ and $s = s_1 = s_2$ (see (3)), contradicting the assumption $s_1 \neq s_2$. If $p_1 > p_2$, $s^* \notin \widetilde{\mathcal{S}}$ and a Helstrom family exists, then $\widetilde{p}_1 = 1$ implies that $p_1 = p = p_2/\widetilde{p}_2$, $s = s_1 = \widetilde{p}_2 s_2 + (1-\widetilde{p}_2)t_2$ and

$$t_2 = \frac{s_1 - \widetilde{p}_2 s_2}{1 - \widetilde{p}_2} = \frac{p_1 s_1 - p_1 \widetilde{p}_2 s_2}{p_1 - p_1 \widetilde{p}_2} = s^* \notin \widetilde{\mathcal{S}} \ ,$$

a contradiction. Thus the first part of the lemma hold. For the second part, if $p_1 > p_2$ and $s^* \in \widetilde{\mathcal{S}}$, then we have $s_1 = (1 - p_2/p_1)s^* + (p_2/p_1)s_2$ by the definition of $s^*$, therefore for any $\mathbf{O} = (1 - e, e) \in \mathcal{O}_2$ we have

$$\begin{aligned} P_{\mathrm{succ}}(\mathbf{O}) &= p_1(1 - e(s_1)) + p_2 e(s_2) \\ &= p_1 - p_1 \left( \left( 1 - \frac{p_2}{p_1} \right) \widetilde{e}(s^*) + \frac{p_2}{p_1} e(s_2) \right) + p_2 e(s_2) \\ &= p_1 - (p_1 - p_2)\widetilde{e}(s^*) \ . \end{aligned}$$

Since $s^* \in \widetilde{\mathcal{S}}$, we have $\widetilde{e}(s^*) \geq 0$, therefore $P_{\mathrm{succ}}(\mathbf{O}) \leq p_1$ for any $\mathbf{O} \in \mathcal{O}_2$. This means that it is a non-generic case. Hence Lemma 5.1 holds. □

Owing to this lemma, in what follows we assume that the condition (4) in Lemma 5.1 is satisfied unless otherwise specified, in order to focus on generic cases. In the following subsections we will prove that a Helstrom family always exists under the assumption (4), that is our main result in this article.

### 5.2 Auxiliary results

In this subsection, for later use we summarize some known facts for topological vector spaces, together with some further properties. Our main reference is the book [18]. See also Sect. 2 for terminology.

First we list the following (special cases of the) facts presented in [18]:

**Theorem 5.1** (Theorem 3.2 in [18, Chap. I]). *Any $n$-dimensional Hausdorff t.v.s. with $n < \infty$ is isomorphic to the $n$-dimensional Euclidean space $\mathbb{R}^n$.*

**Proposition 5.1** (Proposition 3.3 in [18, Chap. I]). *Let $W$ be a t.v.s. If $W'$ is a linear subspace of $W$ that is closed in $W$, and $W''$ is a finite-dimensional linear subspace of $W$, then $W' + W''$ is closed in $W$.*

**Proposition 5.2** (Proposition 3.4 in [18, Chap. I]). *Every linear functional on a finite-dimensional Hausdorff t.v.s. is continuous.*

The next theorem is a variant of Hahn-Banach's Theorem. Here we use the following notion: A real-valued function $g$ on a vector space $W$ is called a *semi-norm* if we have $g(x + y) \leq g(x) + g(y)$ for any $x, y \in W$ and we have $g(\lambda x) = |\lambda| g(x)$ for any $x \in W$ and $\lambda \in \mathbb{R}$. Then we have the following theorem:

**Theorem 5.2** (Theorem 3.2 in [18, Chap. II]). *Let $W$ be a vector space, $g$ a semi-norm on $W$, and $W'$ a linear subspace of $W$. If $f$ is a linear functional on $W'$ such that $|f(x)| \leq g(x)$ for all $x \in W'$, then $f$ extends to a linear functional $\overline{f}$ on $W$ such that $|\overline{f}(x)| \leq g(x)$ for all $x \in W$.*

A subset $C$ of a vector space $W$ is called *circled* if $x \in C$ and $-1 \leq \lambda \leq 1$ imply $\lambda x \in C$; and called *radial* if for any $x \in W$, there exists $\lambda_0 \in \mathbb{R}$ such that $x \in \lambda C$ whenever $|\lambda| \geq |\lambda_0|$. If $C$ is convex, radial and circled, then the *Minkowski functional* (or *gauge*) $g_C : W \to \mathbb{R}$ of $C$ is defined by

$$g_C(x) = \inf\{\lambda > 0 \mid x \in \lambda C\} \text{ for each } x \in W \quad . \tag{5}$$

**Proposition 5.3** (Proposition 1.4 in [18, Chap. II]). *The Minkowski functional $g_C$ of $C$ is a semi-norm on $W$.*

From now, we present the following two properties of our minimal framework (see Theorem 2.1) that are consequences of the above facts:

**Corollary 5.1.** *Every finite-dimensional affine subspace $W$ of the t.v.s. $\widetilde{V}$ is closed in $\widetilde{V}$ and is isomorphic to the Euclidean space $\mathbb{R}^n$ with $n = \dim W$. Hence $\widetilde{\mathcal{S}} \cap W$ is a compact subset of $W$.*

*Proof.* The compactness of $\widetilde{\mathcal{S}} \cap W$ follows from the remaining parts. Since the topology $\mathcal{T}(\widetilde{V})$ of $\widetilde{V}$ is invariant under any translation $x \mapsto x + x_0$, $x_0 \in \widetilde{V}$, we assume without loss of generality that $W$ is a linear subspace of $\widetilde{V}$. Since $\widetilde{V}$ is Hausdorff, the assertion $W \simeq \mathbb{R}^n$ follows from Theorem 5.1; while the null subspace $\{0\}$ of $\widetilde{V}$ is closed in $\widetilde{V}$, therefore $W = \{0\} + W$ is also closed by Proposition 5.1. Hence Corollary 5.1 holds. □

**Corollary 5.2.** *Let $W$ be a finite-dimensional affine subspace of $\widetilde{V}$. Then any affine functional $f$ on $W$ extends to a continuous affine functional $\overline{f}$ on $\widetilde{V}$.*

*Proof.* Fix an element $x_0 \in W$ and put $\alpha = f(x_0)$. Then the linear functional $g : x \mapsto f(x + x_0) - \alpha$ on a finite-dimensional linear subspace $W - x_0$ of $\widetilde{V}$ is continuous by Proposition 5.2 since $\widetilde{V}$ is Hausdorff. Moreover, since $\widetilde{V}$ is l.c., a consequence of Hahn-Banach's Theorem (Theorem D.1) implies that this $g$ extends to a $\overline{g} \in \mathcal{L}_c(\widetilde{V})$. Now the map $\overline{f}$ defined by $\overline{f}(x) = \overline{g}(x - x_0) + \alpha$ is an affine extension of $f$, and $\overline{f}$ is continuous since the translation $x \mapsto x - x_0$ is an isomorphism from $\widetilde{V}$ to itself. Hence Corollary 5.2 holds. □

### 5.3 Candidates of conjugate states for Helstrom families

In this subsection, we investigate the candidates of conjugate states $t_1, t_2$ for Helstrom families. For the purpose, we introduce some further notations. Recall that we have assumed the condition (4). In the case $p_1 = p_2$, let $\mathcal{C}_{\mathrm{weak}}$ be the set of all pairs $(t_1, t_2)$ of distinct $t_1, t_2 \in \widetilde{\mathcal{S}}$ such that the vector $\overrightarrow{t_1 t_2}$ is proportional to $\overrightarrow{s_2 s_1}$ (i.e., $t_2 - t_1 = c(s_1 - s_2)$ for some $c > 0$). On the other hand, in the case $p_1 > p_2$ and $s^* \notin \widetilde{\mathcal{S}}$, let $\mathcal{C}_{\mathrm{weak}}$ be the set of all pairs $(t_1, t_2)$ of distinct $t_1, t_2 \in \widetilde{\mathcal{S}}$ such that $t_2$ lies in the line segment $\overline{t_1 s^*} = \mathrm{Conv}(\{t_1, s^*\})$ between $t_1$ and $s^*$. Note that $\mathcal{C}_{\mathrm{weak}} \neq \emptyset$ since $(s_2, s_1) \in \mathcal{C}_{\mathrm{weak}}$. The next lemma shows that $\mathcal{C}_{\mathrm{weak}}$ is the set of the pairs of conjugate states for weak Helstrom families:

**Lemma 5.2.** *If $(\widetilde{p}_i, s_i; 1 - \widetilde{p}_i, t_i)$, $i = 1, 2$, is a weak Helstrom family, then $(t_1, t_2) \in \mathcal{C}_{\mathrm{weak}}$. Conversely, if $(t_1, t_2) \in \mathcal{C}_{\mathrm{weak}}$, then there exist $0 < \widetilde{p}_i \leq 1$, $i = 1, 2$, such that $(\widetilde{p}_i, s_i; 1 - \widetilde{p}_i, t_i)$, $i = 1, 2$, is a weak Helstrom family.*

*Proof.* First, we consider the case $p_1 = p_2$. Then any weak Helstrom family satisfies $\widetilde{p}_1 = \widetilde{p}_2$, therefore (3) implies that $\widetilde{p}_1 < 1$ (otherwise, we have $s_1 = s = s_2$, contradicting the fact $s_1 \neq s_2$) and $t_2 - t_1 = \widetilde{p}_1(s_1 - s_2)/(1 - \widetilde{p}_1)$. Thus $(t_1, t_2) \in \mathcal{C}_{\mathrm{weak}}$. Conversely, if $(t_1, t_2) \in \mathcal{C}_{\mathrm{weak}}$, then $t_2 - t_1 = c(s_1 - s_2)$ for some $c > 0$, while this $c$ can be written as $c = \widetilde{p}/(1 - \widetilde{p})$ with $0 < \widetilde{p} < 1$. Now it follows that $(\widetilde{p}, s_i; 1 - \widetilde{p}, t_i)$, $i = 1, 2$, is a weak Helstrom family. Thus the lemma holds in this case.

Secondly, we consider the case $p_1 > p_2$ and $s^* \notin \widetilde{\mathcal{S}}$. Then by (3), any weak Helstrom family satisfies $\widetilde{p}_2 = p_2/p = \widetilde{p}_1 p_2/p_1 < \widetilde{p}_1 \leq 1$, therefore

$$t_2 = \frac{\widetilde{p}_1 s_1 - \widetilde{p}_2 s_2 + (1 - \widetilde{p}_1) t_1}{1 - \widetilde{p}_2} = \lambda t_1 + (1 - \lambda) s^* \ , \text{ where } \lambda = \frac{1 - \widetilde{p}_1}{1 - \widetilde{p}_2} \ . \tag{6}$$

Now we have $0 \leq \lambda < 1$ since $\widetilde{p}_2 < \widetilde{p}_1 \leq 1$, therefore $t_2 \in \overline{t_1 s^*}$. Moreover, if $t_1 = t_2$, then (6) implies that $t_1 = s^*$, contradicting $t_1 \in \widetilde{\mathcal{S}}$ and $s^* \notin \widetilde{\mathcal{S}}$. Thus $(t_1, t_2) \in \mathcal{C}_{\mathrm{weak}}$. Conversely, if $(t_1, t_2) \in \mathcal{C}_{\mathrm{weak}}$, then we have $t_2 = \lambda t_1 + (1 - \lambda) s^*$ for some $0 \leq \lambda < 1$, and now $(\widetilde{p}_i, s_i; 1 - \widetilde{p}_i, t_i)$, $i = 1, 2$, is a weak Helstrom family for $\widetilde{p}_1 = (p_1 - p_1\lambda)/(p_1 - p_2\lambda)$ and $\widetilde{p}_2 = \widetilde{p}_1 p_2/p_1$. Hence Lemma 5.2 holds. □

By the lemma and Corollary 4.1, for finding a Helstrom family, it suffices to search a pair $(t_1, t_2) \in \mathcal{C}_{\mathrm{weak}}$ such that $t_1$ and $t_2$ are distinguishable by a virtual effect $\widetilde{e} \in \widetilde{\mathcal{E}}$ (see Definition 4.3 for terminology). The outline to prove the existence of such a pair $(t_1, t_2)$ is the following:

1. Define a function $\ell : \mathcal{C}'_{\mathrm{weak}} \to \mathbb{R}$, where

   $$\mathcal{C}'_{\mathrm{weak}} = \mathcal{C}_{\mathrm{weak}} \cup \{(t, t) \mid t \in \widetilde{\mathcal{S}}\} \subset \widetilde{\mathcal{S}} \times \widetilde{\mathcal{S}} \ ,$$

   such that $\ell \geq 0$ and $\ell(t_1, t_2) = 0$ if and only if $t_1 = t_2$; hence $\ell > 0$ on $\mathcal{C}_{\mathrm{weak}}$.

2. Prove that $\mathcal{C}'_{\mathrm{weak}}$ is closed in $\widetilde{\mathcal{S}} \times \widetilde{\mathcal{S}}$; hence $\mathcal{C}'_{\mathrm{weak}}$ is compact since $\widetilde{\mathcal{S}} \times \widetilde{\mathcal{S}}$ is.

3. Prove that $\ell$ is continuous; hence $\ell$ takes the maximum value at some pair $(t_1, t_2)$ in $\mathcal{C}_{\mathrm{weak}}$ (see the first step).

4. Prove that $t_1$ and $t_2$ are distinguishable.

From now, we proceed the program. In what follows, for a t.v.s. $W$, let $\mathcal{L}(W)$, $\mathcal{L}_c(W)$, $\mathcal{A}(W)$, and $\mathcal{A}_c(W)$ denote, respectively, the sets of linear functionals on $W$, of continuous linear functionals on $W$, of affine functionals on $W$, and of continuous affine functionals on $W$.

For the first step of the program, we define the function $\ell : \mathcal{C}'_{\text{weak}} \to \mathbb{R}$ as follows: In the case $p_1 = p_2$, define $\ell(t_1, t_2)$ by

$$t_2 - t_1 = \ell(t_1, t_2)(s_1 - s_2) \text{ for } (t_1, t_2) \in \mathcal{C}'_{\text{weak}} \ .$$

On the other hand, in the case $p_1 > p_2$ and $s^* \notin \widetilde{\mathcal{S}}$, define $\ell(t_1, t_2)$ by

$$t_2 = \ell(t_1, t_2)s^* + (1 - \ell(t_1, t_2))t_1 \text{ for } (t_1, t_2) \in \mathcal{C}'_{\text{weak}}$$

(thus $0 \leq \ell < 1$; note that $\ell \neq 1$ since $t_2 \neq s^*$). This $\ell$ has the properties specified in the first step. Note that $\ell(t_1, t_2)$ becomes larger if and only if $t_1$ and $t_2$ become "far" from each other in the space $\widetilde{\mathcal{S}}$ (in an intuitive sense; this becomes a strict sense at least in finite-dimensional cases, since in such a case $\widetilde{\mathcal{S}}$ admits the Euclidean metric); hence our program to make the value $\ell(t_1, t_2)$ as large as possible also fits the strategy mentioned in Example 4.1 for decreasing the Helstrom ratio. Namely, in the case $p_1 = p_2$, the definition of $\ell$ intuitively implies that $\ell(t_1, t_2)$ is the "distance" between $t_1$ and $t_2$ normalized as the "distance" between $s_1$ and $s_2$ being 1. On the other hand, in the case $p_1 > p_2$ and $s^* \notin \widetilde{\mathcal{S}}$, the definition of $\ell$ implies that

$$t_1 - t_2 = \frac{\ell(t_1, t_2)}{1 - \ell(t_1, t_2)}(t_2 - s^*) \ ,$$

therefore $\ell(t_1, t_2)/(1 - \ell(t_1, t_2))$, that is increasing for $\ell(t_1, t_2)$, is the "distance" between $t_1$ and $t_2$ normalized as the "distance" between $t_2$ and $s^*$ being 1.

For the second step, we have the following result:

**Lemma 5.3.** *Let* $(t_1, t_2) \in \widetilde{\mathcal{S}} \times \widetilde{\mathcal{S}}$*.*

1. *If* $p_1 = p_2$*, then we have* $(t_1, t_2) \in \mathcal{C}'_{\text{weak}}$ *if and only if* $\widetilde{e}(t_1) \geq \widetilde{e}(t_2)$ *for any* $e \in \mathcal{E}$ *such that* $e(s_1) \leq e(s_2)$*.*
2. *If* $p_1 > p_2$*, then we have* $(t_1, t_2) \in \mathcal{C}'_{\text{weak}}$ *if and only if* $f(t_1) \leq f(t_2) \leq f(s^*)$ *or* $f(t_1) \geq f(t_2) \geq f(s^*)$ *for any* $f \in \mathcal{A}_c(\widetilde{V})$ *such that* $f|_{\widetilde{\mathcal{S}}} \in \widetilde{\mathcal{E}}$*.*

*Proof.* Since the case $t_1 = t_2$ is trivial, we assume from now that $t_1 \neq t_2$.

For the first part, if $(t_1, t_2) \in \mathcal{C}_{\text{weak}}$, then Lemma 5.2 implies that

$$s = \widetilde{p}s_1 + (1 - \widetilde{p})t_1 = \widetilde{p}s_2 + (1 - \widetilde{p})t_2 \text{ for some } 0 < \widetilde{p} < 1 \text{ and } s \in \widetilde{\mathcal{S}}$$

(recall that $s_1 \neq s_2$). Now for any $e \in \mathcal{E}$, we have

$$\widetilde{e}(s) = \widetilde{p}e(s_1) + (1 - \widetilde{p})\widetilde{e}(t_1) = \widetilde{p}e(s_2) + (1 - \widetilde{p})\widetilde{e}(t_2) \ ,$$

therefore $\widetilde{e}(t_1) \geq \widetilde{e}(t_2)$ whenever $e(s_1) \leq e(s_2)$. On the other hand, if $(t_1, t_2) \notin \mathcal{C}'_{\text{weak}}$, then we have either $(t_2, t_1) \in \mathcal{C}_{\text{weak}}$, or $\overrightarrow{t_1 t_2} = t_2 - t_1$ is not parallel to the line $\text{Aff}(\{s_1, s_2\})$ containing $s_1$ and $s_2$. In the former case, we have $e(s_1) < e(s_2)$ for some $e \in \mathcal{E}$ since $\mathcal{S}$ is separated (note that $1 - e \in \mathcal{E}$ and $1 - e(s_1) < 1 - e(s_2)$ whenever $e \in \mathcal{E}$ and $e(s_1) > e(e_2)$), therefore we have $\widetilde{e}(t_2) > \widetilde{e}(t_1)$ in the same way as above. In the latter case, it is easy to show that $f(s_1) = f(s_2)$ and $f(t_1) < f(t_2)$ for an affine functional $f$ on the affine hull of $\{s_1, s_2, t_1, t_2\}$, and Corollary 5.2 implies that this $f$ extends to an $\overline{f} \in \mathcal{A}_c(\widetilde{V})$. Now $\overline{f}(\widetilde{\mathcal{S}})$ is bounded in $\mathbb{R}$ since $\widetilde{\mathcal{S}}$ is compact. Thus by taking $\alpha > 0$ and $\beta \in \mathbb{R}$ appropriately, the continuous affine functional $g = \alpha\overline{f} + \beta$ on $\widetilde{V}$ satisfies that $g(s_1) = g(s_2)$, $g(t_1) < g(t_2)$ and $g(\widetilde{\mathcal{S}}) \subset [0, 1]$, therefore $\widetilde{e} = g|_{\widetilde{\mathcal{S}}}$ is a virtual effect satisfying $e(s_1) = e(s_2)$ and $\widetilde{e}(t_1) < \widetilde{e}(t_2)$. Thus the first part of Lemma 5.3 holds.

For the second part, note that $s^* \notin \widetilde{\mathcal{S}}$ by the assumption (4). The "only if" part is now trivial by the definition of $\mathcal{C}_{\text{weak}}$. To prove the "if" part, assume that $(t_1, t_2) \notin \mathcal{C}'_{\text{weak}}$. Then $t_1 \neq t_2$, and we have either $t_1 \in \overline{t_2 s^*}$, or $\overrightarrow{t_1 t_2}$ is not parallel to the line $\mathrm{Aff}(\{t_1, s^*\})$ (note that $s^* \notin \overline{t_1 t_2}$ since $s^* \notin \widetilde{\mathcal{S}}$). In the former case, since $\widetilde{V}$ is Hausdorff, there exists an $f \in \mathcal{L}_c(\widetilde{V})$ such that $f(t_2) < f(t_1)$. Now since $\widetilde{\mathcal{S}}$ is compact, an appropriate transformation $g = \alpha f + \beta$ with $\alpha, \beta \in \mathbb{R}$ satisfies that $g(\widetilde{\mathcal{S}}) \subset [0,1]$ (hence $g|_{\widetilde{\mathcal{S}}} \in \widetilde{\mathcal{E}}$) and $g(t_2) < g(t_1)$, therefore $g(t_1) < g(s^*)$ since $t \in \overline{t_2 s^*}$ and $t_1 \neq s^*$. On the other hand, in the latter case, we have $f(t_1) = f(s^*) < f(t_2)$ for an affine functional $f$ on the affine hull of $\{t_1, t_2, s^*\}$, and Corollary 5.2 implies that this $f$ extends to an $\overline{f} \in \mathcal{A}_c(\widetilde{V})$. Now $\overline{f}(\widetilde{\mathcal{S}})$ is bounded in $\mathbb{R}$ since $\widetilde{\mathcal{S}}$ is compact. Thus by taking an appropriate affine transformation of $\overline{f}$ in the same way as above, it follows that $g(t_1) = g(s^*) < g(t_2)$ for a $g \in \mathcal{A}_c(\widetilde{V})$ such that $g|_{\widetilde{\mathcal{S}}} \in \widetilde{\mathcal{E}}$. Hence the second part of Lemma 5.3 holds, concluding the proof of Lemma 5.3. □

By this lemma, $\mathcal{C}'_{\mathrm{weak}}$ is closed in $\widetilde{\mathcal{S}} \times \widetilde{\mathcal{S}}$ as desired, since the virtual effect $\widetilde{e} \in \widetilde{\mathcal{E}}$ corresponding to each $e \in \mathcal{E}$ is continuous on $\widetilde{\mathcal{S}}$.

For the third step, we have the following result:

**Lemma 5.4.** *The function $\ell$ on $\mathcal{C}'_{\mathrm{weak}}$ is continuous.*

*Proof.* First, we consider the case $p_1 = p_2$. Fix $e \in \mathcal{E}$ such that $e(s_1) > e(s_2)$ (this is possible since $\mathcal{S}$ is separated), and put $c = (e(s_1) - e(s_2))^{-1} > 0$. For any $(t_1, t_2) \in \mathcal{C}'_{\mathrm{weak}}$, Lemma 5.2 implies that there exists a $\widetilde{p} \in [0,1)$ such that $\widetilde{p} s_1 + (1 - \widetilde{p}) t_1 = \widetilde{p} s_2 + (1 - \widetilde{p}) t_2 \in \widetilde{\mathcal{S}}$. Now we have $\ell(t_1, t_2) = \widetilde{p}/(1 - \widetilde{p})$ and $\widetilde{p} e(s_1) + (1 - \widetilde{p}) \widetilde{e}(t_1) = \widetilde{p} e(s_2) + (1 - \widetilde{p}) \widetilde{e}(t_2)$, therefore

$$\ell(t_1, t_2) = \frac{\widetilde{e}(t_2) - \widetilde{e}(t_1)}{e(s_1) - e(s_2)} = c(\widetilde{e}(t_2) - \widetilde{e}(t_1)) \enspace .$$

This implies that $\ell$ is continuous, since $\widetilde{e} \in \widetilde{\mathcal{E}}$ is continuous.

Secondly, we consider the case that $p_1 > p_2$ and $s^* \notin \widetilde{\mathcal{S}}$. Let $\mathcal{F}$ be the set of all $f \in \mathcal{A}_c(\widetilde{V})$ such that $f|_{\widetilde{\mathcal{S}}} \in \widetilde{\mathcal{E}}$. Now for each $f \in \mathcal{F}$, put

$$A_f = \{(t_1, t_2) \in \widetilde{\mathcal{S}} \times \widetilde{\mathcal{S}} \mid f(t_1) \neq f(s^*)\} \subset \widetilde{\mathcal{S}} \times \widetilde{\mathcal{S}}$$

and define a function $g_f : A_f \to \mathbb{R}$ by

$$g_f(t_1, t_2) = \frac{f(t_2) - f(t_1)}{f(s^*) - f(t_1)} \text{ for } (t_1, t_2) \in A_f \enspace .$$

Since $f$ is continuous, $A_f$ is open in $\widetilde{\mathcal{S}} \times \widetilde{\mathcal{S}}$ and $g_f$ is continuous. Moreover, we have $\ell(t_1, t_2) = g_f(t_1, t_2)$ for any $(t_1, t_2) \in \mathcal{C}'_{\mathrm{weak}} \cap A_f$ by the definition of $\ell$. Now we show that

$$\ell^{-1}(U) = \bigcup_{f \in \mathcal{F}} (\mathcal{C}'_{\mathrm{weak}} \cap {g_f}^{-1}(U)) \text{ for any open subset } U \subset \mathbb{R} \enspace .$$

Once this is proven, $\ell^{-1}(U)$ is open in $\mathcal{C}'_{\mathrm{weak}}$ since each ${g_f}^{-1}(U) \subset A_f$ is an open subset of $\widetilde{\mathcal{S}} \times \widetilde{\mathcal{S}}$ (recall that $A_f$ is open in $\widetilde{\mathcal{S}} \times \widetilde{\mathcal{S}}$), therefore the continuity of $\ell$ follows. Since $\ell$ and $g_f$ agree on $\mathcal{C}'_{\mathrm{weak}} \cap A_f$ as above, the inclusion $\supset$ holds immediately. For the other inclusion, let $(t_1, t_2) \in \mathcal{C}'_{\mathrm{weak}}$ such that $\ell(t_1, t_2) \in U$. Let $W$ denote the line $\mathrm{Aff}(\{s_1, s_2\})$. Now if $t_1 \notin W$, then an argument similar to Lemma 5.3 (based on Corollary 5.2) implies existence of an $f \in \mathcal{F}$ such that $f$ is constant on $W$ and $f(t_1) \neq f(s_1)$, hence $f(s^*) = f(s_1) \neq f(t_1)$ (note that $s^* \in W$). On the other hand, suppose that $t_1 \in W$. Since $\widetilde{V}$ is Hausdorff, there exists an $f \in \mathcal{A}_c(\widetilde{V})$ such that $f(s_1) \neq f(s_2)$. Now by a similar argument as above, this $f$ can be chosen from $\mathcal{F}$. Since the four points $s_1$, $s_2$, $s^*$, and $t_1$ are all collinear and $t_1 \neq s^*$, the fact $f(s_1) \neq f(s_2)$ implies that $f(s^*) \neq f(t_1)$. Thus $(t_1, t_2) \in A_f$ in any case, while $\ell$ and $g_f$ agree on $\mathcal{C}'_{\mathrm{weak}} \cap A_f$, therefore $g_f(t_1, t_2) = \ell(t_1, t_2) \in U$ by the above argument. Hence the inclusion $\subset$ follows, therefore Lemma 5.4 holds. □

For the final part, let $\mathcal{C}$ be the subset of $\mathcal{C}'_{\text{weak}}$ that consists of all pairs in $\mathcal{C}'_{\text{weak}}$ at which $\ell$ takes the maximum value:

$$\mathcal{C} = \{(t_1, t_2) \in \mathcal{C}'_{\text{weak}} \mid \ell(t_1, t_2) = \max_{(t'_1, t'_2) \in \mathcal{C}'_{\text{weak}}} \ell(t'_1, t'_2)\} \ .$$

Note that $\emptyset \neq \mathcal{C} \subset \mathcal{C}_{\text{weak}}$ by the above argument. From now, we show that for any $(t_1, t_2) \in \mathcal{C}_{\text{weak}}$, $t_1$ and $t_2$ are distinguishable if and only if $(t_1, t_2) \in \mathcal{C}$; in particular, a Helstrom family exists. First, one direction of this assertion is proven as follows:

**Proposition 5.4.** *If $(t_1, t_2) \in \mathcal{C}_{\text{weak}}$, and $t_1$ and $t_2$ are distinguishable, then $(t_1, t_2) \in \mathcal{C}$. Hence the pair of conjugate states $t_1, t_2$ in any Helstrom family belongs to $\mathcal{C}$.*

*Proof.* The latter part is derived from the combination of the former part, Lemma 5.2, Lemma 5.1, and Corollary 4.2. To prove the former part, assume contrary that $t_1$ and $t_2$ in $\widetilde{\mathcal{S}}$ are distinguishable by a virtual effect $\widetilde{e} \in \widetilde{\mathcal{E}}$ and $(t_1, t_2) \in \mathcal{C}_{\text{weak}}$ but $\ell(t_1, t_2) < \ell(t'_1, t'_2)$ for some $(t'_1, t'_2) \in \mathcal{C}_{\text{weak}}$. Since $\text{Aff}(\widetilde{\mathcal{S}}) = \widetilde{V}$, this $\widetilde{e}$ extends to an $f \in \mathcal{A}(\widetilde{V})$. Let $W = \text{Aff}(\{t_1, t_2, t'_1, t'_2\})$. Then by Corollary 5.1, $W$ is isomorphic to a Euclidean space $\mathbb{R}^n$ with $n = \dim W$ and $\mathcal{S}' = \widetilde{\mathcal{S}} \cap W$ is a compact convex subset of $W$. Since $(t_1, t_2), (t'_1, t'_2) \in \mathcal{C}_{\text{weak}}$, we have $n \leq 2$ by the definition of $\mathcal{C}_{\text{weak}}$. Now $H_1 = W \cap f^{-1}(1)$ and $H_2 = W \cap f^{-1}(0)$ are parallel supporting hyperplanes of $\mathcal{S}'$ in $W$ at $t_1$ and at $t_2$, respectively, and $\mathcal{S}'$ lies between $H_1$ and $H_2$. Note that $t_1, t_2, t'_1, t'_2 \in \mathcal{S}'$.

Now in the case $p_1 = p_2$, $\overrightarrow{t'_1 t'_2}$ is parallel to $\overrightarrow{t_1 t_2}$ since $(t_1, t_2), (t'_1, t'_2) \in \mathcal{C}_{\text{weak}}$. Thus it is geometrically obvious that $|t'_1 t'_2| \leq |t_1 t_2|$ (where $|xy|$ denotes the distance between $x$ and $y$ in the Euclidean metric on $\mathbb{R}^n$), since two intersecting points of the line $\text{Aff}(\{t'_1, t'_2\})$ with $H_1$ and with $H_2$, respectively, and $t_1$ and $t_2$ form a parallelogram (see Fig. 2(a)). This contradicts the assumption $\ell(t'_1, t'_2) > \ell(t_1, t_2)$.

On the other hand, we consider the case that $p_1 > p_2$ and $s^* \notin \widetilde{\mathcal{S}}$. Note that $s^* \in W$ since $(t_1, t_2) \in \mathcal{C}_{\text{weak}}$. Then the assumption $\ell(t_1, t_2) < \ell(t'_1, t'_2)$ implies that $|t'_2 t'_1|/|s^* t'_2| > |t_2 t_1|/|s^* t_2|$; in particular, neither $t'_1$ nor $t'_2$ lies on the line segment $\overline{t_1 t_2}$. Let $v_1$ and $v_2$ be the intersecting points of the line $\text{Aff}(\{t'_1, t'_2\})$ with $H_1$ and with $H_2$, respectively (see Fig. 2(b)). Then we have

$$\frac{|v_2 v_1|}{|s^* v_2|} \geq \frac{|t'_2 t'_1|}{|s^* t'_2|} > \frac{|t_2 t_1|}{|s^* t_2|} \ .$$

However, since $H_1$ and $H_2$ are parallel, two triangles $\triangle s^* v_1 t_1$ and $\triangle s^* v_2 t_2$ are similar, therefore we have $|v_2 v_1|/|s^* v_2| = |t_2 t_1|/|s^* t_2|$, a contradiction.

Thus a contradiction occurs in both cases. Hence Proposition 5.4 holds. □

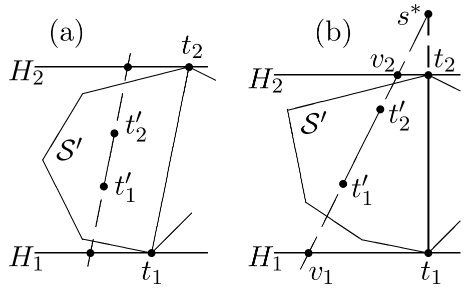

Figure 2: The cases (a) $p_1 = p_2$ and (b) $p_1 > p_2$, $s^* \notin \widetilde{\mathcal{S}}$ in Proposition 5.4

Now we are in a position to state our main theorem in this article, that will be proven in the next subsection:

**Theorem 5.3.** *If $(t_1, t_2) \in \mathcal{C}$, then $t_1$ and $t_2$ are distinguishable. Hence, by the above argument, a Helstrom family always exists under the assumption* (4)*.*

Before starting the proof we notice the following: Once Theorem 5.3 is proven, the hypothesis "and a Helstrom family exists" in the first part of Lemma 5.1 becomes redundant, therefore the following simple criterion for generic cases in two-state discrimination problems will be obtained that improves Lemma 5.1:

**Theorem 5.4.** *Under the assumption $p_1 \geq p_2$, the condition* (4) *is necessary and sufficient for the case to be generic.*

In particular, an equiprobable case $p_1 = p_2 = 1/2$ is always a generic case, therefore in such a case we can always make a correct guess with probability strictly higher than $1/2$ by using an appropriate observable.

We also mention another nontrivial consequence of Theorem 5.3 that shows interesting relations between optimal success probabilities for equiprobable two-state discrimination problems and Gudder's metric functions on the state space (cf., Remark 2.3):

*Remark* 5.1. First, we translate the definition of Gudder's metric function [9] on compact state spaces to our framework with not necessarily compact real state space $\mathcal{S}$. For $s'_1, s'_2 \in \mathcal{S}$, define $d(s'_1, s'_2)$ to be the infimum of $0 < \lambda \leq 1/2$ such that

$$\lambda t_1 + (1-\lambda)s'_1 = \lambda t_2 + (1-\lambda)s'_2 \text{ for some } t_1, t_2 \in \widetilde{\mathcal{S}}$$

(note that $\lambda = 1/2$, $t_1 = s'_2$ and $t_2 = s'_1$ always satisfy this condition). This function $d$ is a metric on $\mathcal{S}$, and this definition coincides with Gudder's original definition in the case $\widetilde{\mathcal{S}} = \mathcal{S}$ (i.e., when $\mathcal{S}$ is compact). Now the above condition is equivalent to that $(1-\lambda, s'_i; \lambda, t_i)$, $i = 1, 2$, is a weak Helstrom family for states $s'_1, s'_2$ and a priori probabilities $p_i = 1/2$, with Helstrom ratio given by $p = 1/(2-2\lambda)$. Thus minimizing $\lambda$ is equivalent to minimizing $p$, and Theorem 5.3 implies that the infimum $d(s'_1, s'_2)$ of such $\lambda$ is attained by some Helstrom family, with Helstrom ratio $p = P_{\mathrm{succ}}(s'_1, s'_2)$ where $P_{\mathrm{succ}}(s'_1, s'_2)$ denotes the optimal success probability for discriminating $s'_1$ and $s'_2$ in the equiprobable case. Thus we have a nontrivial relation

$$d(s'_1, s'_2) = 1 - \frac{1}{2P_{\mathrm{succ}}(s'_1, s'_2)} \text{ for any } s'_1, s'_2 \in \mathcal{S} \ . \tag{7}$$

In particular, it follows that the function of $s'_1, s'_2$ in the right-hand side is a metric on $\mathcal{S}$. It seems infeasible to derive the fact just from the intuitive meaning of "optimal success probability of state discrimination".

On the other hand, Gudder also defined another metric function on the same state space, called the "intrinsic metric", by using the former metric function $d$ as a building block. According to Gudder's definition, we put

$$\widetilde{d}(s'_1, s'_2) = \frac{d(s'_1, s'_2)}{1 - d(s'_1, s'_2)} \text{ for } s'_1, s'_2 \in \mathcal{S} \ .$$

The concrete structure of the above metric $d$ implies that $\widetilde{d}$ is indeed a metric function and $0 \leq \widetilde{d} \leq 1$. Moreover, it follows from (7) that

$$\widetilde{d}(s'_1, s'_2) = 2P_{\mathrm{succ}}(s'_1, s'_2) - 1 \text{ for } s'_1, s'_2 \in \mathcal{S} \ . \tag{8}$$

This shows an operational meaning of Gudder's intrinsic metric that has not been pointed out in the literature. Moreover, by comparing (8) to the well-known formula $P_{\mathrm{succ}}(\rho_1, \rho_2) =$

$1/2+D(\rho_1,\rho_2)/2$ for *quantum* states $\rho_1, \rho_2$, where $D(\rho_1,\rho_2)$ denotes the trace distance, Gudder's intrinsic metric coincides with the trace distance for quantum cases. Hence we have obtained an operationally natural generalization of the trace distance to general probabilistic theories.

Moreover, it is in fact possible to define the "trace distance" in general probabilistic theories directly through the classical trace distance:

$$D(s'_1, s'_2) = \sup_{\mathbf{O}=(e_i)_i\in\mathcal{O}} D_c(e_i(s'_1), e_i(s'_2)) \text{ for } s'_1, s'_2 \in \mathcal{S} \ , \tag{9}$$

where $D_c(p_i, q_i)$ denotes the classical trace distance ($L_1$ distance or Kolmogorov distance) [15] between probability distributions $p_i$ and $q_i$:

$$D_c(p_i, q_i) = \frac{1}{2}\sum_i |p_i - q_i| \ ,$$

and $\mathcal{O} = \bigcup_{N\in\mathbb{N}} \mathcal{O}_N$ denotes the set of all discrete observables. (Note that the argument below shows that the supremum in (9) is always attained by some observable and it can be chosen from two-valued observables.) Since the classical trace distance is the maximal difference of probabilities between $p_i$ and $q_i$ among all events $S$, i.e., $D_c(p_i, q_i) = \max_S |p(S) - q(S)| = \max_S |\sum_{i\in S} p_i - \sum_{i\in S} q_i|$, it is considered as an operationally natural distance between probability distributions. In order to distinguish states $s'_1$ and $s'_2$ in general probabilistic theories, what one can do best is to find the best observable $\mathbf{O} = (e_i)_i \in \mathcal{O}$ for catching the difference between $s'_1$ and $s'_2$ by comparing the probability distributions $e_i(s'_1)$ and $e_i(s'_2)$. Thus we are lead to the definition (9) of the distance between states; namely, $D(s'_1, s'_2)$ has the same operational meaning as Kolmogorov distance that is optimal among all observables. From now, we show that Gudder's intrinsic metric (8) is in fact the same as our trace distance (9).

For the purpose, first we show that in our trace distance, it suffices to consider just two-valued observables $\mathbf{O} = (e_i)_i \in \mathcal{O}_2$, namely:

$$D(s'_1, s'_2) = \sup_{\mathbf{O}=(e_i)_i\in\mathcal{O}_2} D_c(e_i(s'_1), e_i(s'_2)) \ . \tag{10}$$

(Now the supremum is attained by some observable due to the compactness of $\mathcal{O}_2$ and the continuity of $D_c(e_i(s'_1), e_i(s'_2))$; see the proof of Theorem 3.1.) To prove (10), note that one can associate to any $\mathbf{O} = (e_i)_i \in \mathcal{O}$ a two-valued observable $(e'_+, e'_-) \in \mathcal{O}_2$ with $e'_+ = \sum_{i\in M_+} e_i$ and $e'_- = 1 - e_+$, where $M_+ = \{i \mid e_i(s'_1) \geq e_i(s'_2)\}$. By the definition, we have $D_c(e_i(s'_1), e_i(s'_2)) = D_c(e'_\pm(s'_1), e'_\pm(s'_2))$. This implies that the right-hand side of (10) is greater than or equal to the right-hand side of (9), while the opposite inequality holds obviously (since $\mathcal{O}_2 \subset \mathcal{O}$). Hence (10) holds. Note that this argument also provides another simple expression of our trace distance $D(s'_1, s'_2)$:

$$D(s'_1, s'_2) = \sup_{e\in\mathcal{E}} \left[e(s'_1) - e(s'_2)\right] \ , \tag{11}$$

where the supremum is again attained by some effect due to the compactness of $\mathcal{E}$ (see the proof of Theorem 3.1).

Now it is not difficult to see that Gudder's intrinsic metric (8) is indeed the same as our trace distance (9): To see this, just observe that for $s'_1, s'_2 \in \mathcal{S}$ with a priori probabilities $p_1 = p_2 = 1/2$, we have from (1) and (2)

$$P_{\text{succ}}(s'_1, s'_2) = \frac{1}{2}(1 + \sup_{e\in\mathcal{E}} \left[e(s'_1) - e(s'_2)\right]) \ .$$

Substituting it into (8) and using (11), we obtain the desired relation:

$$\widetilde{d}(s'_1, s'_2) = D(s'_1, s'_2) \ . \tag{12}$$

The equivalence (12) provides simple proofs for several properties of $\widetilde{d}(s_1', s_2')$ originally shown by Gudder [9]. For instance, since the classical trace distance $D_c(p_i, q_i)$ is well known to be a metric, so is our trace distance $D(s_1', s_2')$ by the definition, therefore $\widetilde{d}(s_1', s_2')$ is indeed a metric as well (for positiveness of $D(s_1', s_2')$ with $s_1' \neq s_2'$ we needed the fact that the state space is separated). We also consider another important property, the *monotonicity* of $\widetilde{d}(s_1', s_2')$:

*Theorem* 5.5 (Gudder [9]). *For any state $s_1', s_2' \in \mathcal{S}$ and any affine map $F : \mathcal{S} \to \mathcal{S}$, we have*

$$\widetilde{d}(F(s_1'), F(s_2')) \leq \widetilde{d}(s_1', s_2') \ .$$

Now this fact is an easy consequence of the equivalence (12) and the fact that affine maps are closed under composition. Namely, for any observable $(e_i)_i \in \mathcal{O}$, by putting $f_i = e_i \circ F : \mathcal{S} \to [0, 1]$ we have

$$D_c(e_i(F(s_1')), e_i(F(s_2'))) = D_c(f_i(s_1'), f_i(s_2')) \ . \tag{13}$$

Now $(f_i)_i$ is also an observable, therefore the supremum of the left-hand side of (13) over $(e_i)_i \in \mathcal{O}$ does not exceed the supremum of the right-hand side of (13) over *all* observables $(f_i)_i$. This implies the monotonicity of $D(s_1', s_2')$, hence of $\widetilde{d}(s_1', s_2')$. (We note that the quantity in the right-hand side of (11) was also investigated in [5] in slightly different context; for instance, it was shown to be a metric, and the monotonicity was also proven there.)

Summarizing, we have shown that Gudder's intrinsic metric has two operational meanings; one is directly given through the classical trace distance (12); another is given by the optimal success probability to discriminate two states under a uniform distribution (8).

*Remark* 5.2. As an application of Gudder's intrinsic metric, or the trace distance defined above, we have a simple (qualitative) version of information disturbance theorem in general probabilistic theories. Before giving the theorem, we clarify the meaning of some terminology. We say that a state $s$ is a *pure state* if $s$ is an extremal point of the state space. We say that two states are *indistinguishable* if these are not distinguishable in the sense of Definition 4.3. Then the above-mentioned theorem is the following:

*Theorem* 5.6. *In any general probabilistic theory, any attempt to distinguish two indistinguishable pure states causes a disturbance.*

This theorem is a generalization of the well-known corresponding theorem in quantum theory (see e.g., Proposition 12.18 in [15]) to arbitrary general probabilistic theories. It is known that a general probabilistic theory is non-classical if and only if there exist indistinguishable pure states [1]. Hence one can conclude that the information disturbance property inevitably holds for any non-classical general probabilistic theory, not only for quantum theory.

Before presenting the proof, notice that any dynamics on $\mathcal{S}$ should be described by an affine map $F : \mathcal{S} \to \mathcal{S}$ in order to preserve the probabilistic mixture, while the composition of state spaces $\mathcal{S}_1$ and $\mathcal{S}_2$ is given by a tensor product $\mathcal{S}_1 \otimes \mathcal{S}_2$ (see [1] and references therein).

*Theorem 5.6.* Let $s_1, s_2 \in \mathcal{S}$ be two indistinguishable pure states (thus $1 > P_{\mathrm{succ}}(s_1, s_2)$). Let $s_i \otimes s_0$ $(i = 1, 2)$ be the initial states on $\mathcal{S} \otimes \mathcal{S}'$, where $s_0 \in \mathcal{S}'$ is any fixed state to which the information of $s_1$ or $s_2$ is transferred. Assume contrary that one can extract information with which one distinguishes $s_1$ and $s_2$ without causing any disturbance. More precisely, we assume that there exists an information transfer machine described by an affine map $F : \mathcal{S} \otimes \mathcal{S}' \to \mathcal{S} \otimes \mathcal{S}'$ such that the reduced states of $F(s_i \otimes s_0)$ to the first system $\mathcal{S}$ remains to be $s_i$ (i.e., causing no disturbance) while the reduced states of $F(s_1 \otimes s_0)$ and $F(s_2 \otimes s_0)$ to the second system $\mathcal{S}'$ are distinct (i.e., enabling one to extract some information to distinguish $s_1$ and $s_2$). Now it is easy to show that if a reduced state is in pure state, then the whole state should be a product state by showing that there exist no correlations between an arbitrary pair of observables (or effects). Therefore, we have

$$F(s_1 \otimes s_0) = s_1 \otimes t_1 \ , \ \ F(s_2 \otimes s_0) = s_2 \otimes t_2 \ ,$$

with $t_1 \neq t_2 \in \mathcal{S}'$. Using the machine $F$ $N$ times, one obtains an affine transformation $\widetilde{F}$ on $\mathcal{S} \otimes \mathcal{S}'^{\otimes N}$ such that

$$\widetilde{F}(s_i \otimes (s_0^{\otimes N})) = s_i \otimes t_i^{\otimes N} \ .$$

Physically, this means that one obtains an arbitrary large number of ensembles for (distinct) state $t_1$ or $t_2$, and thereby can distinguish them with success probability arbitrarily close to 1. In other words, the optimal success probability to distinguish $\widetilde{F}(s_1 \otimes (s_0^{\otimes N}))$ and $\widetilde{F}(s_1 \otimes (s_2^{\otimes N}))$ can be exponentially close to 1 with respect to $N$ (to see this formally, use Chernoff bound [7] for instance). On the other hand, we have

$$1 > P_{\mathrm{succ}}(s_1, s_2) = P_{\mathrm{succ}}(s_1 \otimes s_0^{\otimes N}, s_2 \otimes s_0^{\otimes N}) \geq P_{\mathrm{succ}}(\widetilde{F}(s_1 \otimes s_0^{\otimes N}), \widetilde{F}(s_2 \otimes s_0^{\otimes N}))$$

for any $N$, where the last inequality follows from Theorem 5.5 and (8). This is a contradiction, since the last term converges to 1 when $N \to \infty$ as mentioned above. Hence the proof of Theorem 5.6 is concluded. □

### 5.4 Proof of Theorem 5.3

In this subsection, we give a proof of Theorem 5.3, namely we prove that $t_1$ and $t_2$ in $\widetilde{\mathcal{S}}$ are distinguishable if $(t_1, t_2) \in \mathcal{C}$ (see Definition 4.3 for terminology).

First, we would like to reduce our argument to the special case $t_2 = -t_1$. For the purpose, let $v_0 = (t_1 + t_2)/2 \in \widetilde{\mathcal{S}}$ and put $C = \widetilde{\mathcal{S}} - v_0$, that is also a convex subset of $\widetilde{V}$. Moreover, put $\overline{t_1} = t_1 - v_0$ and $\overline{t_2} = t_2 - v_0$. Then we have $\overline{t_1}, \overline{t_2} \in C$ and $\overline{t_2} = -\overline{t_1}$. Note that $\overline{t_1} \neq \overline{t_2}$ since $t_1 \neq t_2$.

The outline of our proof is the following. First, note that the existence of an $\widetilde{e} \in \widetilde{\mathcal{E}}$ such that $\widetilde{e}(t_1) = 1$ and $\widetilde{e}(t_2) = 0$ (that is nothing but our goal) is obvious if $\widetilde{V}$ coincides with the 1-dimensional linear subspace $W'$ spanned by $\overline{t_1}$ (hence by $\overline{t_2}$). To construct such an $\widetilde{e}$ in more general case, we would like to extend a nonzero linear functional $f$ on $W'$ (note that $f$ is continuous on $W'$ and $f(C \cap W')$ is bounded in $\mathbb{R}$) to a continuous linear functional $\overline{f}$ on $\widetilde{V}$ such that $\overline{f}(C)$ is bounded in $\mathbb{R}$. Then it will be shown that the restriction of an appropriate affine transformation $h = \alpha\overline{f} + \beta$ of $\overline{f}$ ($\alpha, \beta \in \mathbb{R}$) to $\widetilde{\mathcal{S}}$ is the desired virtual effect $\widetilde{e}$. To construct such an extension $\overline{f}$ of $f$, first we use Theorem 5.2 to obtain an extension $f'$ of $f$ to $W = \widetilde{V}$ (not yet necessarily continuous) such that $f'(C)$ is bounded in $\mathbb{R}$, and then we further modify the functional $f'$ by using Theorem D.1 to obtain $\overline{f}$.

To perform the program, we start with the linear functional $f$ on the 1-dimensional subspace $W'$ such that $f(\lambda\overline{t_1}) = \lambda$ for each $\lambda \in \mathbb{R}$, therefore $f(\overline{t_1}) = 1$ and $f(\overline{t_2}) = -1$. To apply Theorem 5.2, we would like to take an appropriate semi-norm $g$ on $\widetilde{V}$, more precisely, the Minkowski functional $g_{\widetilde{C}}$ of a certain subset $\widetilde{C}$ of $\widetilde{V}$ (see Proposition 5.3). From now, we define the subset $\widetilde{C}$. Note that the convex subset $C$ of $\widetilde{V}$ contains the origin of $\widetilde{V}$, therefore we have $\lambda x \in C$ for any $x \in C$ and $0 \leq \lambda \leq 1$. Thus the subset $\pm C = C \cup -C$ of $\widetilde{V}$ is circled (see Sect. 5.2 for terminology). Now define $\widetilde{C}$ to be the convex hull $\mathrm{Conv}(\pm C)$ of $\pm C$, which is also a circled subset of $\widetilde{V}$. By the convexity of $C$, any element $v$ of $\widetilde{C}$ can be written as $v = \lambda x - \lambda' x'$ with $x, x' \in C$, $\lambda, \lambda' \geq 0$ and $\lambda + \lambda' = 1$. This subset $\widetilde{C}$ has the following property:

**Lemma 5.5.** *$\widetilde{C}$ is a radial subset of $\widetilde{V}$ (see Sect. 5.2 for terminology).*

*Proof.* Let $W_0$ be the set of all $v \in \widetilde{V}$ such that $v \in \lambda\widetilde{C}$ for some $\lambda > 0$. Then $W_0$ contains $\widetilde{C}$, hence $C$. Moreover, if $v \in W_0$, $\lambda_0 > 0$ and $v \in \lambda_0\widetilde{C}$, then we have $v \in \lambda\widetilde{C}$ whenever $|\lambda| \geq \lambda_0$ since $\widetilde{C}$ is circled. Thus $\widetilde{C}$ is radial if $W_0 = \widetilde{V}$. To prove $W_0 = \widetilde{V}$, it suffices to show that $W_0$ is a linear subspace of $\widetilde{V}$. Indeed, once this is proven, $W_0 + v_0$ will be an affine subspace of $\widetilde{V}$ containing $\widetilde{\mathcal{S}}$ (recall that $W_0 \supset \widetilde{C} = \widetilde{\mathcal{S}} - v_0$), therefore $W_0 + v_0 = \widetilde{V}$ (hence $W_0 = \widetilde{V}$) since $\mathrm{Aff}(\widetilde{\mathcal{S}}) = \widetilde{V}$.

Let $v_1, v_2 \in W_0$. Then for each $i$, we have $v_i \in \lambda_i x_i$ for some $\lambda_i > 0$ and $x_i \in \widetilde{C}$. Moreover, let $\mu_1, \mu_2 \in \mathbb{R}$, and write $\mu_i = \varepsilon_i \nu_i$ with $\varepsilon_i \in \{\pm 1\}$ and $\nu_i \geq 0$ for each $i$. We show that $\mu_1 v_1 + \mu_2 v_2 \in W_0$; since this is obvious when $\mu_1 = \mu_2 = 0$, we assume from now that $\nu_1 > 0$ or $\nu_2 > 0$. Then by putting $x'_i = \varepsilon_i x_i \in \widetilde{C}$ for each $i$ (note that $\widetilde{C}$ is circled), we have

$$\mu_1 v_1 + \mu_2 v_2 = \lambda_1\nu_1 x'_1 + \lambda_2\nu_2 x'_2 = (\lambda_1\nu_1 + \lambda_2\nu_2)\frac{\lambda_1\nu_1 x'_1 + \lambda_2\nu_2 x'_2}{\lambda_1\nu_1 + \lambda_2\nu_2} \enspace ,$$

therefore $\mu_1 v_1 + \mu_2 v_2 \in (\lambda_1\nu_1 + \lambda_2\nu_2)\widetilde{C}$ by the convexity of $\widetilde{C}$. Hence we have $\mu_1 v_1 + \mu_2 v_2 \in W_0$, therefore Lemma 5.5 holds. □

Owing to the above properties of $\widetilde{C}$, we define the semi-norm $g$ to be the Minkowski functional $g_{\widetilde{C}}$ of $\widetilde{C}$ (see Proposition 5.3). Note that $g(v) \leq 1$ for any $v \in \widetilde{C}$ by the definition of $g = g_{\widetilde{C}}$.

From now, to apply Theorem 5.2, we show that $|f(v)| \leq g(v)$ for any $v \in W'$. Since $W'$ is 1-dimensional and $g$ is a semi-norm, it suffices to show that $g(\overline{t_1}) = 1 = f(\overline{t_1})$. This is proven in the following lemma:

**Lemma 5.6.** *We have $g(\overline{t_1}) = 1$.*

*Proof.* First, note that $g(\overline{t_1}) \leq 1$ since $\overline{t_1} \in \widetilde{C}$. We show that $g(\overline{t_1}) \geq 1$, or equivalently, there does not exist an element $v \in \widetilde{C}$ and $c > 1$ such that $v = c\overline{t_1}$. Assume contrary that such a pair $(v, c)$ exists. As mentioned before, this $v$ is of the form $v = \lambda x - (1-\lambda)x'$ with $x, x' \in C$ and $0 \leq \lambda \leq 1$, therefore $\lambda x - (1-\lambda)x' = c\overline{t_1} = -c\overline{t_2}$. Moreover, by the definition of $C$, we have $x = s - v_0$ and $x' = s' - v_0$ for some $s, s' \in \widetilde{\mathcal{S}}$, therefore

$$v = \lambda s - (1-\lambda)s' + (1-2\lambda)v_0 = c\overline{t_1} = -c\overline{t_2} \enspace .$$

Note also that $t_2 - t_1 = \overline{t_2} - \overline{t_1} = 2\overline{t_2} = -2\overline{t_1}$. From now, we show that we can construct a pair $(t'_1, t'_2) \in \mathcal{C}'_{\mathrm{weak}}$ (by using the convexity of $\widetilde{\mathcal{S}}$) such that $\ell(t'_1, t'_2) > \ell(t_1, t_2)$, contradicting the assumption $(t_1, t_2) \in \mathcal{C}$.

First we consider the case that $p_1 = p_2$. If $\lambda \leq 1/2$, then we have

$$(1-\lambda)s' - \lambda s - (1-2\lambda)t_1 = -v - (1-2\lambda)\overline{t_1} = (c+1-2\lambda)\overline{t_2} \enspace ,$$

therefore $s' - s'' = \alpha(t_2 - t_1)$, where $s'' = (\lambda s + (1-2\lambda)t_1)/(1-\lambda) \in \widetilde{\mathcal{S}}$ (note that $\widetilde{\mathcal{S}}$ is convex) and $\alpha = (c+1-2\lambda)/(2-2\lambda)$. Since $c > 1$, we have $\alpha > 1$, therefore $(s'', s') \in \mathcal{C}'_{\mathrm{weak}}$ and $\ell(s'', s') = \alpha\ell(t_1, t_2) > \ell(t_1, t_2)$, as desired. Similarly, if $\lambda \geq 1/2$, then we have

$$(2\lambda - 1)t_2 + (1-\lambda)s' - \lambda s = -v + (2\lambda - 1)\overline{t_2} = (c + 2\lambda - 1)\overline{t_2} \enspace ,$$

therefore $s'' - s = \alpha(t_2 - t_1)$ where $s'' = (2 - \lambda^{-1})t_2 + (\lambda^{-1} - 1)s' \in \widetilde{\mathcal{S}}$ and $\alpha = (c + 2\lambda - 1)/(2\lambda) > 1$. Thus we have $(s, s'') \in \mathcal{C}'_{\mathrm{weak}}$ and $\ell(s, s'') > \ell(t_1, t_2)$, as desired.

Secondly, we consider the case that $p_1 > p_2$ and $s^* \not\in \widetilde{\mathcal{S}}$. Put $\ell = \ell(t_1, t_2)$ for simplicity. Note that $0 < \ell < 1$ and

$$\ell s^* = t_2 - (1-\ell)t_1 = 2v_0 - (2-\ell)t_1 = (2-\ell)t_2 - (2-2\ell)v_0 \enspace ,$$

while

$$v = \lambda s - (1-\lambda)s' + (1-2\lambda)v_0 = ct_1 - cv_0 = cv_0 - ct_2 \ .$$

Put $\mu = (2-\ell)(2\lambda-1) + c\ell$. If $\mu \geq 0$, then the above relations imply that

$$\lambda(2-2\ell)s + \ell(c+2\lambda-1)s^* = (1-\lambda)(2-2\ell)s' + \mu t_2 \ .$$

Now the coefficients of $s$, $s^*$, $s'$, and $t_2$ in this equality are all nonnegative, and the sums of the two coefficients in the left-hand side and in the right-hand side, respectively, are positive and equal to each other; namely,

$$\lambda(2-2\ell) + \ell(c+2\lambda-1) = (1-\lambda)(2-2\ell) + \mu = c\ell + 2\lambda - \ell > 0 \ .$$

Thus by the convexity of $\widetilde{\mathcal{S}}$, we have $(1-\alpha)s + \alpha s^* = s''$ for some $s'' \in \widetilde{\mathcal{S}}$, where

$$\alpha = \frac{\ell(c+2\lambda-1)}{c\ell+2\lambda-\ell} = 1 - \frac{2\lambda(1-\ell)}{c\ell+2\lambda-\ell} \in (\ell, 1]$$

(note that $0 < \ell < 1$ and $c > 1$). Thus we have $(s, s'') \in \mathcal{C}'_{\text{weak}}$ and $\ell(s, s'') = \alpha > \ell$, as desired. Similarly, if $\mu < 0$, then we have

$$2\lambda s + |\mu| t_1 + \ell(c+1-2\lambda)s^* = (2-2\lambda)s' \ .$$

Since $c > 1$, all the four coefficients in this equality are nonnegative, and the sum of the three coefficients in the left-hand side is equal to the coefficient $2 - 2\lambda > 0$ in the right-hand side; namely,

$$2\lambda + |\mu| + \ell(c+1-2\lambda) = 2 - 2\lambda > 0 \ .$$

Thus by the convexity of $\widetilde{\mathcal{S}}$, we have $(1-\alpha)s'' + \alpha s^* = s'$ for some $s'' \in \widetilde{\mathcal{S}}$, where $\alpha = \ell(c+1-2\lambda)/(2-2\lambda) \in (\ell, 1]$ (note that $\ell > 0$ and $c > 1$). Thus we have $(s'', s') \in \mathcal{C}'_{\text{weak}}$ and $\ell(s'', s') = \alpha > \ell$, as desired.

Hence our claim holds in all cases, therefore Lemma 5.6 holds. □

Thus by Theorem 5.2, the functional $f$ on $W'$ extends to an $f' \in \mathcal{L}(\widetilde{V})$ such that $|f'(v)| \leq g(v)$ for any $v \in \widetilde{V}$. Since $f'|_{W'} = f$, we have $f'(\overline{t_1}) = 1$, $f'(\overline{t_2}) = -1$ and $|f'(x)| \leq g(x) \leq 1$ for any $x \in \widetilde{C}$, therefore $f'(C) \subset [-1, 1]$. By putting $\alpha = f'(v_0)$, it follows that

$$f'(t_1) = \alpha + 1 \ , \ \ f'(t_2) = \alpha - 1 \ , \ \ f'(\widetilde{\mathcal{S}}) \subset [\alpha-1, \alpha+1] \ ,$$

therefore the restriction of $f'$ to $V$ is continuous. Our desired virtual effect $\widetilde{e}$ can be constructed directly from this $f'$ if $f'$ is also continuous on $\widetilde{V}$; however, this is not guaranteed in general.

Thus, instead, by using Theorem D.1, we take a continuous linear functional $\overline{f}$ on $\widetilde{V}$ such that $\overline{f}|_V = f'|_V$. Note that $\overline{f}(\mathcal{S}) \subset [\alpha-1, \alpha+1]$ since $\mathcal{S} \subset \widetilde{\mathcal{S}} \cap V$, therefore we have $\overline{f}(\widetilde{\mathcal{S}}) \subset [\alpha-1, \alpha+1]$ since $\widetilde{\mathcal{S}} = \mathrm{cl}_{\widetilde{V}}(\mathcal{S})$. From now, we show that $\overline{f}(t_1) = \alpha + 1$ and $\overline{f}(t_2) = \alpha - 1$. First, we consider the case $p_1 = p_2$. Then we have $t_1 - t_2 = c(s_2 - s_1)$ with $c = \ell(t_1, t_2) > 0$, while $s_2 - s_1 \in V$ since $s_1, s_2 \in \mathcal{S}$, therefore

$$\overline{f}(t_2 - t_1) = c\overline{f}(s_1 - s_2) = cf'(s_1 - s_2) = f'(t_2 - t_1) = -2 \ .$$

Since $\overline{f}(\widetilde{\mathcal{S}}) \subset [\alpha-1, \alpha+1]$ as mentioned above, we have

$$\alpha - 1 \leq \overline{f}(t_2) = \overline{f}(t_1) - 2 \leq \alpha + 1 - 2 = \alpha - 1 \ ,$$

therefore $\overline{f}(t_2) = \alpha - 1$ and $\overline{f}(t_1) = \overline{f}(t_2) + 2 = \alpha + 1$. Secondly, we consider the case $p_1 > p_2$ and $s^* \notin \widetilde{\mathcal{S}}$. Now $t_2 = cs^* + (1-c)t_1$ with $0 < c = \ell(t_1, t_2) < 1$, while $s^* \in V$, therefore

$$\overline{f}(t_2) - (1-c)\overline{f}(t_1) = c\overline{f}(s^*) = cf'(s^*) = f'(t_2) - (1-c)f'(t_1) \ .$$

Now we have $\overline{f}(t_1) \leq \alpha + 1 = f'(t_1)$, therefore $\overline{f}(t_2) \leq f'(t_2) = \alpha - 1$ since $1 - c > 0$. Thus we have $\overline{f}(t_2) = \alpha - 1$ since $\overline{f}(t_2) \geq \alpha - 1$, therefore $\overline{f}(t_1) = f'(t_1) = \alpha + 1$. Hence we have $\overline{f}(t_1) = \alpha + 1$ and $\overline{f}(t_2) = \alpha - 1$ in any case.

Finally, by the above properties, the affine functional $h = (\overline{f} + 1 - \alpha)/2$ on $\widetilde{V}$ is continuous and satisfies that $h(t_1) = 1$, $h(t_2) = 0$ and $h(\widetilde{\mathcal{S}}) \subset [0,1]$. This implies that $\widetilde{e} = h|_{\widetilde{\mathcal{S}}}$ is a virtual effect that distinguishes $t_1$ and $t_2$.

Hence the proof of Theorem 5.3 is concluded.

**Acknowledgments.** The authors would like to thank Dr. Manabu Hagiwara, Dr. Kentaro Imafuku, and Professor Hideki Imai, for their significant comments. A part of this work was supported by Grant-in-Aid for Young Scientists (B) (20700017), The Ministry of Education, Culture, Sports, Science and Technology (MEXT).

# Appendix: Proof of Theorem 2.1

In the appendix, we give a proof of Theorem 2.1. In what follows, For any convex structure $C$, let $\mathcal{A}^b_{C'}(C)$ be the set of all $f \in \mathcal{A}(C)$ bounded on a subset $C'$ of $C$. Moreover, for any convex subset $C$ of a t.v.s., let $\mathcal{A}_c(C)$ denote the set of all continuous $f \in \mathcal{A}(C)$.

## A Construction of $\mathcal{S}$ and $V$

First, we describe construction of a vector space $V$ and its convex subset $\mathcal{S}$ such that $\mathcal{S}$ is isomorphic to the separated convex structure $\overline{\mathcal{S}_0}$ and $V = \mathrm{Aff}(\mathcal{S})$. Here we abuse the notations $\mathcal{S}$ and $V$ though these $\mathcal{S}$ and $V$ are in fact not necessarily the same as (but isomorphic to) $\mathcal{S}$ and $V$ in Theorem 2.1, respectively. Although our argument is essentially the standard one (cf., [1, 9, 11, 13, 14, 16]), we give the argument here for the sake of completeness.

Our argument is the following. In what follows, let $\mathcal{L}(W)$ denote the set of all linear functionals on a vector space $W$; and for any convex structure $C$, let $\mathcal{A}(C)$ denote the set of all affine functionals on $C$. Then the set $\mathcal{A}(\overline{\mathcal{S}_0})$ forms a vector space with natural addition and scalar multiplication, therefore its dual space $\mathcal{A}(\overline{\mathcal{S}_0})^* = \mathcal{L}(\mathcal{A}(\overline{\mathcal{S}_0}))$ is also a vector space. We define an "evaluation map" $\mathsf{ev}_s : \mathcal{A}(\overline{\mathcal{S}_0}) \to \mathbb{R}$ for each $s \in \overline{\mathcal{S}_0}$ by $\mathsf{ev}_s(f) = f(s)$ for $f \in \mathcal{A}(\overline{\mathcal{S}_0})$. Then a straightforward argument shows that $\mathsf{ev}_s \in \mathcal{A}(\overline{\mathcal{S}_0})^*$ for every $s \in \overline{\mathcal{S}_0}$, and the map $\psi : \overline{\mathcal{S}_0} \to \mathcal{A}(\overline{\mathcal{S}_0})^*$, $\psi(s) = \mathsf{ev}_s$, is a homomorphism of convex structures, i.e., $\psi(\langle \lambda, \mu; s, t \rangle) = \lambda\psi(s) + \mu\psi(t)$ for any $s, t \in \overline{\mathcal{S}_0}$. The fact that $\overline{\mathcal{S}_0}$ is separated (Lemma 2.1) implies that $\psi$ is injective. Moreover, by fixing an element $v \in \psi(\overline{\mathcal{S}_0})$, the map $\varphi : \overline{\mathcal{S}_0} \to \mathcal{A}(\overline{\mathcal{S}_0})^*$, $\varphi(s) = \psi(s) - v$, is also an injective homomorphism of convex structures. Thus $\mathcal{S} = \varphi(\overline{\mathcal{S}_0})$ is a convex subset of the vector space $\mathcal{A}(\overline{\mathcal{S}_0})^*$ containing the origin of $\mathcal{A}(\overline{\mathcal{S}_0})^*$. Now $V = \mathrm{Aff}(\mathcal{S})$ is a linear subspace of $\mathcal{A}(\overline{\mathcal{S}_0})^*$. Thus $\mathcal{S}$ and $V$ are obtained.

## B Topologies on $\mathcal{S}$ and $V$

Secondly, we give the definition of topologies on $V$ and $\mathcal{S}$. In what follows, for any vector space $W$, let $\mathcal{L}^b_C(W)$ denote the set of all $f \in \mathcal{L}(W)$ bounded on a given subset $C$ of $W$. For any t.v.s. $W$, let $\mathcal{L}_c(W)$ denote the set of all continuous $f \in \mathcal{L}(W)$. For a convex subset $C$ of a vector space $W$ and a subset $\mathcal{F}$ of $\mathcal{A}(C)$, let $\sigma(C,\mathcal{F})$ denote the weakest topology on $C$ to make every $f \in \mathcal{F}$ continuous. For a topology $\mathcal{T}$ on a space $X$ and a subset $Y$ of $X$, let $\mathcal{T}|_Y$ denote the relative topology on $Y$ induced by $\mathcal{T}$. For two topologies $\mathcal{T}$ and $\mathcal{T}'$ on the same set $X$, we write $\mathcal{T} \subset \mathcal{T}'$ to signify that $\mathcal{T}'$ is stronger than or equal to $\mathcal{T}$ (i.e., every $\mathcal{T}$-open subset of $X$ is $\mathcal{T}'$-open). Moreover, let $\mathcal{E}$ denote the set of all $e \in \mathcal{A}(S)$ such that $e(\mathcal{S}) \subset [0,1]$.

Now we define the topology $\mathcal{T}(V)$ on $V$ by

$$\mathcal{T}(V) = \sigma(V, \mathcal{L}^b_{\mathcal{S}}(V)) \enspace .$$

This topology makes $V$ a l.c.t.v.s. (see e.g., [18, Chap. II, Sect. 5]). Moreover, this $V$ satisfies the following property:

**Lemma B.1.** *The t.v.s. $V$ is Hausdorff.*

*Proof.* First, since $\mathcal{S}$ is convex, an elementary argument shows that the affine hull $\mathrm{Aff}(\mathcal{S}) = V$ of $\mathcal{S}$ consists of all elements of the form $\lambda s - \lambda' s'$ with $s, s' \in \mathcal{S}$, $\lambda \geq 1$ and $\lambda - \lambda' = 1$. Let $v = \lambda s - \lambda' s'$ and $v' = \mu t - \mu' t'$ be distinct elements of $V$ written in the above form. Now put

$$p = \frac{\lambda - 1}{\lambda + \mu - 1} \enspace , \enspace q = \frac{\mu - 1}{\lambda + \mu - 1} \enspace , \enspace r = \frac{1}{\lambda + \mu - 1} \enspace ,$$

therefore $p, q \geq 0$, $r > 0$ and $p + q + r = 1$. Moreover, put

$$w = ps' + qt' + rv \enspace , \enspace w' = ps' + qt' + rv' \enspace .$$

Then $w \neq w'$ since $v \neq v'$ and $r > 0$, while we have

$$w = r\lambda s + (p - r\lambda')s' + qt' = (1-q)s + qt' \in \mathcal{S}$$

since $\mathcal{S}$ is convex, and similarly $w' \in \mathcal{S}$. Since $\mathcal{S} \simeq \overline{\mathcal{S}_0}$ is separated by Lemma 2.1, there exists an $e \in \mathcal{E}$ such that $e(w) \neq e(w')$. Now by the definitions of $w$ and $w'$, the affine extension $f$ of $e$ to $V$ satisfies $f \in \mathcal{L}^b_{\mathcal{S}}(V)$ and $f(v) \neq f(v')$. Thus $V$ is Hausdorff with respect to $\sigma(V, \mathcal{L}^b_{\mathcal{S}}(V))$. Hence Lemma B.1 holds. □

On the other hand, the induced topology on $\mathcal{S}$ satisfies the following:

**Lemma B.2.** *Two topologies $\mathcal{T}(V)|_{\mathcal{S}}$ and $\sigma(\mathcal{S},\mathcal{E})$ on $\mathcal{S}$ coincide.*

*Proof.* In the proof, put $\mathcal{T} = \mathcal{T}(V) = \sigma(V, \mathcal{L}^b_{\mathcal{S}}(V))$. First, we show that each $e \in \mathcal{E}$ is $(\mathcal{T}|_{\mathcal{S}})$-continuous. Since $\mathrm{Aff}(\mathcal{S}) = V$, this $e$ extends to an affine functional $f$ on $V$ such that $f(\mathcal{S})$ is bounded, therefore $f + \alpha \in \mathcal{L}^b_{\mathcal{S}}(V)$ for some $\alpha \in \mathbb{R}$. Thus $f + \alpha$ is $\mathcal{T}$-continuous by the definition of $\mathcal{T}$, therefore $f$ is also $\mathcal{T}$-continuous and $e = f|_{\mathcal{S}}$ is $(\mathcal{T}|_{\mathcal{S}})$-continuous as desired. This implies that $\sigma(\mathcal{S},\mathcal{E}) \subset \mathcal{T}|_{\mathcal{S}}$.

Now it suffices to show that each $(\mathcal{T}|_{\mathcal{S}})$-open subset $U$ of $\mathcal{S}$ is $\sigma(\mathcal{S},\mathcal{E})$-open. Take a $\mathcal{T}$-open subset $U'$ of $V$ such that $U = U' \cap \mathcal{S}$. Then for each $s \in U \subset U'$, by the definition of $\mathcal{T}$, there exist a finite number of $f_i \in \mathcal{L}^b_{\mathcal{S}}(V)$ and the same number of open subsets $W_i \subset \mathbb{R}$ such that $s \in \bigcap_i f_i^{-1}(W_i) \subset U'$. Since $s \in \mathcal{S}$, we have $s \in \bigcap_i(\mathcal{S} \cap f_i^{-1}(W_i)) \subset U$, therefore it suffices to show that each subset $\mathcal{S} \cap f_i^{-1}(W_i)$ of $\mathcal{S}$ is $\sigma(\mathcal{S},\mathcal{E})$-open. Since $f_i(\mathcal{S})$ is bounded, there exist $\alpha_i, \beta_i \in \mathbb{R}$ such that $\alpha_i \neq 0$ and the functional $g_i = \alpha_i f_i + \beta_i$ satisfies $g_i(\mathcal{S}) \subset [0,1]$, therefore $e_i = g_i|_{\mathcal{S}} \in \mathcal{E}$. Moreover, we have $f_i^{-1}(W_i) = g_i^{-1}(\alpha_i W_i + \beta_i)$ and $W_i' = \alpha_i W_i + \beta_i$ is also an open subset of $\mathbb{R}$. Thus $\mathcal{S} \cap f_i^{-1}(W_i) = \mathcal{S} \cap g_i^{-1}(W_i') = e_i^{-1}(W_i')$, that is $\sigma(\mathcal{S},\mathcal{E})$-open by the definition of $\sigma(\mathcal{S},\mathcal{E})$. Hence Lemma B.2 holds. □

# C The Completions of $\mathcal{S}$ and $V$

To proceed the proof of Theorem 2.1 further, we recall the following notion: The *completion* of a uniform space $X$ is a complete uniform space $\widetilde{X}$ such that $X$ is a dense subspace of $\widetilde{X}$. (See e.g., [4, Chap. II] or [18] for properties of uniform spaces). The completion $\widetilde{X}$ of such a space $X$ always exists, and $\widetilde{X}$ is Hausdorff if and only if $X$ is Hausdorff. Since any t.v.s. is a uniform space (see e.g., Proposition 1.4 in [18, Chap. I]), the completion $\widetilde{V}$ of the Hausdorff t.v.s. $V$ exists in the above sense. Moreover, this $\widetilde{V}$ also admits a structure of a t.v.s., and now $\widetilde{V}$ is a complete Hausdorff t.v.s. and $V$ is a topological vector subspace of $\widetilde{V}$ (with the induced topology equal to $\sigma(V, \mathcal{L}_{\mathcal{S}}^{b}(V))$) that is dense in $\widetilde{V}$ (see e.g., Proposition 1.5 in [18, Chap. I]). Here we use the conventional notation $\widetilde{V}$ for the completion of $V$, though it is not necessarily the same as (but is closely related to) the $\widetilde{V}$ in Theorem 2.1.

Since $\mathcal{S}$ is convex, the closure $\widetilde{\mathcal{S}} = \mathrm{cl}_{\widetilde{V}}(\mathcal{S})$ of $\mathcal{S}$ in $\widetilde{V}$ is also convex in $\widetilde{V}$ (see e.g., Proposition 1.2 in [18, Chap. II]). Again, note that this $\widetilde{\mathcal{S}}$ does not necessarily coincide with (but is closely related to) the $\widetilde{\mathcal{S}}$ in Theorem 2.1. Now the closed subset $\widetilde{\mathcal{S}}$ of the complete t.v.s. $\widetilde{V}$ is also complete (as a uniform subspace), therefore $\widetilde{\mathcal{S}}$ is the completion of $\mathcal{S}$ (as a uniform subspace of $V$) since $\mathcal{S}$ is dense in $\widetilde{\mathcal{S}}$. We would like to show that $\widetilde{\mathcal{S}}$ is compact; we give a lemma for the purpose. Here we use the following terminology. A subset $B$ of a t.v.s. $W$ is called *bounded* if for any 0-neighborhood (i.e., neighborhood of the origin) $U$ of $W$, there exists a $\lambda \in \mathbb{R}$ such that $B \subset \lambda U$. Then we have the following:

**Lemma C.1.** *The convex subset $\mathcal{S}$ of $V$ is bounded in $V$.*

*Proof.* By the definition of the topology on $V$, each 0-neighborhood $U$ of $V$ contains an open 0-neighborhood of the form $\bigcap_i f_i^{-1}(U'_i)$ with finitely many $f_i \in \mathcal{L}_{\mathcal{S}}^{b}(V)$ and the same number of open subsets $U'_i$ of $\mathbb{R}$ containing 0. Since each $f_i(\mathcal{S}) \subset \mathbb{R}$ is bounded, there is a $\lambda > 0$ such that $f_i(\mathcal{S}) \subset \lambda U'_i$ for every $i$. Thus $\mathcal{S} \subset \lambda f_i^{-1}(U'_i)$ for every $i$, therefore $\mathcal{S} \subset \lambda U$. Hence the lemma holds. □

Now note that the topology $\mathcal{T}(V) = \sigma(V, \mathcal{L}_{\mathcal{S}}^{b}(V))$ of $V$ is a weak topology, i.e., it coincides with $\sigma(V, \mathcal{L}_c(V))$ where continuity of each $f \in \mathcal{L}_c(V)$ is with respect to $\mathcal{T}(V)$ (namely, every member of $\mathcal{L}_{\mathcal{S}}^{b}(V)$ is continuous with respect to $\sigma(V, \mathcal{L}_c(V))$ and every member of $\mathcal{L}_c(V)$ is continuous with respect to $\mathcal{T}(V)$). Since $\mathcal{S} \subset V$ is bounded by Lemma C.1, and $V$ is l.c., it follows that $\mathcal{S}$ is *precompact*, i.e., the completion $\widetilde{\mathcal{S}}$ of $\mathcal{S}$ is compact (see e.g., Corollary 2 of Proposition 5.5 in [18, Chapter IV]). The current situation is summarized as follows:

- $\mathcal{S} \simeq \overline{\mathcal{S}_0}$ is a convex subset of a l.c. Hausdorff t.v.s. $V$ containing the origin, with $\mathrm{Aff}(\mathcal{S}) = V$, such that the induced topology on $\mathcal{S}$ is $\sigma(\mathcal{S}, \mathcal{E})$;
- the topology $\mathcal{T}(V)$ of $V$ is $\sigma(V, \mathcal{L}_{\mathcal{S}}^{b}(V)) = \sigma(V, \mathcal{L}_c(V))$;
- $\widetilde{V}$ is a complete Hausdorff t.v.s. containing $V$ as a dense topological vector subspace;
- $\widetilde{\mathcal{S}} = \mathrm{cl}_{\widetilde{V}}(\mathcal{S})$ is the completion of $\mathcal{S}$ that is compact and convex.

# D Existence of the Objects in Theorem 2.1

From now, we modify the above objects to obtain the objects in Theorem 2.1. In what follows, for a t.v.s. $W$, let $\sigma(W)$ denote the weak topology $\sigma(W, \mathcal{L}_c(W))$ on $W$. The following facts will be used in our argument:

**Proposition D.1** (Corollary 2 of Theorem 4.1 in [18, Chap. IV])**.** *Let $W$ be a l.c.t.v.s. with topology $\mathcal{T}$, $W'$ a vector subspace of $W$, and $\overline{W} = W/W'$ the quotient space. Then the weak topology $\sigma(W')$ on $W'$ with respect to $\mathcal{T}|_{W'}$ coincides with $\sigma(W)|_{W'}$, and the weak topology $\sigma(\overline{W})$ on $\overline{W}$ with respect to the quotient topology induced by $\mathcal{T}$ is the quotient topology induced by $\sigma(W)$.*

**Theorem D.1** (Theorem 4.2 in [18, Chap. II])**.** *Let $W$ be a l.c.t.v.s., $W'$ a vector subspace of $W$, and $f \in \mathcal{L}_c(W')$. Then $f$ extends to an $\overline{f} \in \mathcal{L}_c(W)$.*

Note that the weak topology $\sigma(\widetilde{V})$ on $\widetilde{V}$ with respect to the original topology $\mathcal{T}$ of $\widetilde{V}$ is weaker than or equal to $\mathcal{T}$, therefore $\widetilde{\mathcal{S}}$ is also compact with respect to $\sigma(\widetilde{V})$. Now we have the following property:

**Lemma D.1.** *We have $\sigma(\widetilde{V})|_V = \mathcal{T}(V) = \sigma(V, \mathcal{L}^b_{\mathcal{S}}(V))$.*

*Proof.* Note that $\sigma(\widetilde{V})|_V \subset \sigma(V, \mathcal{L}^b_{\mathcal{S}}(V))$ since $\mathcal{T}|_V = \mathcal{T}(V)$ by the definition of $\widetilde{V}$. Thus it suffices to show that each $f \in \mathcal{L}^b_{\mathcal{S}}(V)$ is continuous with respect to $\sigma(\widetilde{V})|_V$. Now this $f$ is $(\mathcal{T}|_V)$-continuous since $\mathcal{T}|_V = \mathcal{T}(V)$, therefore Theorem D.1 implies that $f$ extends to a $\mathcal{T}$-continuous $g \in \mathcal{L}(\widetilde{V})$. This $g$ is also $\sigma(\widetilde{V})$-continuous by the definition of $\sigma(\widetilde{V})$, therefore $f = g|_V$ is continuous with respect to $\sigma(\widetilde{V})|_V$, as desired. Hence the lemma holds. □

In what follows, continuity of a map from $\widetilde{V}$ is considered with respect to $\sigma(\widetilde{V})$ instead of $\mathcal{T}$ unless otherwise specified. Let $\widetilde{V}_0$ denote the intersection of the kernels $\ker(f)$ of all $f \in \mathcal{L}_c(\widetilde{V})$. Let $\pi$ denote the quotient map $\widetilde{V} \to \widetilde{V}/\widetilde{V}_0$, and let $\widetilde{\mathcal{T}} = \pi(\sigma(\widetilde{V}))$ denote the quotient topology on $\pi(\widetilde{V})$ induced by $\sigma(\widetilde{V})$. Note that for any $f \in \mathcal{L}_c(\widetilde{V})$, there exists a unique $\overline{f} \in \mathcal{L}_c(\pi(\widetilde{V}))$ such that $f = \overline{f} \circ \pi$, and any element of $\mathcal{L}_c(\pi(\widetilde{V}))$ is obtained in this manner. Thus by Proposition D.1, the topology $\widetilde{\mathcal{T}}$ of $\pi(\widetilde{V})$ is a weak topology and coincides with $\sigma(\pi(\widetilde{V}), \mathcal{F})$ where $\mathcal{F} = \{\overline{f} \mid f \in \mathcal{L}_c(\widetilde{V})\}$, therefore $\pi(\widetilde{V})$ is a l.c.t.v.s. that is Hausdorff by the definition of $\pi(\widetilde{V})$. Note that $\pi(V)$ is a linear subspace of $\pi(\widetilde{V})$ and $\pi(\mathcal{S})$ is convex in $\pi(V)$. Similarly, $\pi(\widetilde{\mathcal{S}})$ is also convex in $\pi(\widetilde{V})$, and $\pi(\widetilde{\mathcal{S}})$ is compact since $\widetilde{\mathcal{S}}$ is compact and $\pi$ is continuous. On the other hand, since $\sigma(\widetilde{V}) \subset \mathcal{T}$, $V$ is $\mathcal{T}$-dense in $\widetilde{V}$ and $\mathcal{S}$ is $(\mathcal{T}|_{\widetilde{\mathcal{S}}})$-dense in $\widetilde{\mathcal{S}}$, it follows that $V$ is also $\sigma(\widetilde{V})$-dense in $\widetilde{V}$ and $\mathcal{S}$ is also $(\sigma(\widetilde{V})|_{\widetilde{\mathcal{S}}})$-dense in $\widetilde{\mathcal{S}}$, therefore $\pi(V)$ is dense in $\pi(\widetilde{V})$ and $\pi(\mathcal{S})$ is dense in $\pi(\widetilde{\mathcal{S}})$ since $\pi$ is continuous. Moreover, we have the following two properties:

**Lemma D.2.** *We have $\widetilde{\mathcal{T}}|_{\pi(V)} = \sigma(\pi(V), \mathcal{L}^b_{\pi(\mathcal{S})}(\pi(V)))$.*

*Proof.* Since $\widetilde{\mathcal{T}}|_{\pi(V)}$ is a weak topology by Proposition D.1, it suffices to show that an $f \in \mathcal{L}(\pi(V))$ is $(\widetilde{\mathcal{T}}|_{\pi(V)})$-continuous if and only if $f \in \mathcal{L}^b_{\pi(\mathcal{S})}(\pi(V))$. First, let $f \in \mathcal{L}^b_{\pi(\mathcal{S})}(\pi(V))$. Then $f \circ \pi|_V \in \mathcal{L}^b_{\mathcal{S}}(V)$, therefore $f \circ \pi|_V \in \mathcal{L}_c(V)$ by the definition of the topology of $V$. By Lemma D.1, $f \circ \pi|_V$ is also $(\sigma(\widetilde{V})|_V)$-continuous. Thus Theorem D.1 implies that $f \circ \pi|_V$ extends to a $g \in \mathcal{L}_c(\widetilde{V})$. Take the $\overline{g} \in \mathcal{L}_c(\pi(\widetilde{V}))$ corresponding to $g$. Then we have $\overline{g}(\pi(v)) = g(v) = f(\pi(v))$ for any $v \in V$, therefore $\overline{g}|_{\pi(V)} = f$. Thus $f$ is $(\widetilde{\mathcal{T}}|_{\pi(V)})$-continuous.

Secondly, let $f \in \mathcal{L}(\pi(V))$ that is $(\widetilde{\mathcal{T}}|_{\pi(V)})$-continuous. Then by Theorem D.1, this $f$ extends to a $g \in \mathcal{L}_c(\pi(\widetilde{V}))$. Now $g \circ \pi \in \mathcal{L}_c(\widetilde{V})$, therefore $B = g \circ \pi(\widetilde{\mathcal{S}})$ is bounded in $\mathbb{R}$ since $\widetilde{\mathcal{S}}$ is compact. Moreover, we have $f(\pi(s)) = g(\pi(s)) \in B$ for each $s \in \mathcal{S}$, therefore $f(\pi(\mathcal{S})) \subset B$ is also bounded in $\mathbb{R}$. Thus we have $f \in \mathcal{L}^b_{\pi(\mathcal{S})}(\pi(V))$. Hence Lemma D.2 holds. □

**Lemma D.3.** *$\pi|_V$ is a bijection from $V$ to $\pi(V)$.*

*Proof.* Let $v$ and $v'$ be distinct elements of $V$. Then, since $V$ is Hausdorff by Lemma B.1 and the topology of $V$ is a weak topology, there exists an $f \in \mathcal{L}_c(V)$ such that $f(v) \neq f(v')$. Now Lemma D.1 and Theorem D.1 imply that this $f$ extends to a $g \in \mathcal{L}_c(\widetilde{V})$, and we have $g(v) \neq g(v')$. Thus $v - v' \notin \widetilde{V}_0$ and $\pi(v) \neq \pi(v')$. Hence the lemma holds. □

By Lemma D.2, Lemma D.3, and the definition of $\mathcal{T}(V)$, the map $\pi|_V$ is an isomorphism of t.v.s. from $V$ to $\pi(V)$. Moreover, $\pi|_{\mathcal{S}} : \mathcal{S} \to \pi(\mathcal{S})$ is also an isomorphism of convex structures. The current situation is summarized as follows:

- $\pi(\widetilde{V})$ is a l.c. Hausdorff t.v.s. with a weak topology;
- $\pi(V)$ is a topological vector subspace of $\pi(\widetilde{V})$, with induced topology equal to $\sigma(\pi(V), \mathcal{L}^b_{\pi(\mathcal{S})}(\pi(V)))$, that is dense in $\pi(\widetilde{V})$;
- $\pi(\mathcal{S}) \simeq \overline{\mathcal{S}_0}$ is a convex subset of $\pi(V)$ that contains the origin of $\pi(V)$ and satisfies $\mathrm{Aff}(\pi(\mathcal{S})) = \pi(V)$, with the relative topology $\sigma(\pi(\mathcal{S}), \mathcal{E}(\pi(\mathcal{S})))$;
- $\pi(\widetilde{\mathcal{S}})$ is the closure of $\pi(\mathcal{S})$ in $\pi(\widetilde{V})$ that is convex and compact.

Note that the above objects $\pi(\mathcal{S})$, $\pi(V)$, $\pi(\widetilde{\mathcal{S}})$, and $\pi(\widetilde{V})$ will be the desired objects in Theorem 2.1 if the affine hull of $\pi(\widetilde{\mathcal{S}})$ coincides with $\pi(\widetilde{V})$. However, this is not necessarily guaranteed in general. Instead, we take a linear subspace $W = \mathrm{Aff}(\pi(\widetilde{\mathcal{S}}))$ of $\pi(\widetilde{V})$ (note that $\pi(\widetilde{\mathcal{S}})$ contains the origin of $\pi(\widetilde{V})$). Then $W$ is also a l.c. Hausdorff t.v.s., and the topology of $W$ is also a weak topology by Proposition D.1. This $W$ contains $\pi(V)$ since $\pi(V) = \mathrm{Aff}(\pi(\mathcal{S}))$, and $\pi(V)$ is dense in $W$ since it is dense in $\pi(\widetilde{V})$. On the other hand, $\pi(\widetilde{\mathcal{S}})$ is also the compact closure of $\pi(\mathcal{S})$ in $W$ since $\pi(\widetilde{\mathcal{S}}) \subset W$. Moreover, by taking the completion $X$ of the Hausdorff uniform space $\pi(\widetilde{\mathcal{S}})$, the compact subset $\pi(\widetilde{\mathcal{S}})$ of the Hausdorff space $X$ is closed in $X$, therefore $X = \mathrm{cl}_X(\pi(\widetilde{\mathcal{S}})) = \pi(\widetilde{\mathcal{S}})$ and $\pi(\widetilde{\mathcal{S}})$ itself is complete. Thus the objects $\pi(\mathcal{S})$, $\pi(V)$, $\pi(\widetilde{\mathcal{S}})$, and $W$ play the roles of $\mathcal{S}$, $V$, $\widetilde{\mathcal{S}}$, and $\widetilde{V}$ in Theorem 2.1, respectively. Hence the existence of the objects in Theorem 2.1 is proven.

# E Uniqueness of the Objects in Theorem 2.1

Finally, we prove the uniqueness of the objects in Theorem 2.1 (in the sense specified in the statement). Let $(\mathcal{S}, V, \widetilde{\mathcal{S}}, \widetilde{V})$ and $(\mathcal{S}', V', \widetilde{\mathcal{S}}', \widetilde{V}')$ be two collections of the objects as in the statement. First, since $\mathcal{S} \simeq \overline{\mathcal{S}_0} \simeq \mathcal{S}'$, there exists an affine isomorphism $f : \mathcal{S} \to \mathcal{S}'$. Since $V = \mathrm{Aff}(\mathcal{S})$ and $V' = \mathrm{Aff}(\mathcal{S}')$, this $f$ extends to an affine isomorphism $V \to V'$, denoted also by $f$ (thus $f(\mathcal{S}) = \mathcal{S}'$). Now note that the topology $\mathcal{T}(V)$ of $V$ is also the weakest topology to make every *affine* functional $g$ on $V$, such that $g(\mathcal{S})$ is bounded in $\mathbb{R}$, a continuous map. The same also holds for $V'$. Moreover, for each affine functional $g$ on $V$, $g(\mathcal{S})$ is bounded if and only if $g \circ f^{-1}(\mathcal{S}')$ is bounded. Thus it follows from the above properties of $\mathcal{T}(V)$ and $\mathcal{T}(V')$ that the affine isomorphism $f : V \to V'$ is also a homeomorphism of topological spaces.

From now, we show that this $f : V \to V'$ extends to the map $\widetilde{V} \to \widetilde{V}'$ specified in Theorem 2.1. For the purpose, take the completions $W$ and $W'$ of $\widetilde{V}$ and of $\widetilde{V}'$, respectively (cf., Appendix C). Then $W$ is also a Hausdorff t.v.s. and contains $\widetilde{V}$ (hence $V$) as a dense topological vector subspace. The same also holds for $W'$ and $\widetilde{V}'$. Since $W$ and $W'$ are complete, $V$ is dense in $W$, and $V'$ is dense in $W'$, it follows that the above homeomorphism $f : V \to V'$ extends to a homeomorphism $W \to W'$, denoted also by $f$. Now we have the following:

**Lemma E.1.** *The above map $f : W \to W'$ is also an affine isomorphism.*

*Proof.* It suffices to show that $f$ preserves the convex combination of two elements. Let $\lambda, \mu \geq 0$ such that $\lambda + \mu = 1$. Then for each $v, v' \in V$, we have $\lambda f(v) + \mu f(v') = f(\lambda v + \mu v')$ since $f|_V : V \to V'$ is affine. This implies that the two maps $g_1(v, v') = \lambda f(v) + \mu f(v')$ and $g_2(v, v') = f(\lambda v + \mu v')$ from $V \times V$ to $W'$ coincide with each other. Since $V \times V$ is dense in $W \times W$ and $W'$ is complete, the continuous map $g_1 = g_2 : V \times V \to W'$ has a unique continuous extension $W \times W \to W'$. On the other hand, both $\overline{g_1}(w, w') = \lambda f(w) + \mu f(w')$ and $\overline{g_2}(w, w') = f(\lambda w + \mu w')$ are continuous maps from $W \times W$ to $W'$ and satisfy that $\overline{g_1}|_{V \times V} = g_1$ and $\overline{g_2}|_{V \times V} = g_2$. This implies that $\overline{g_1} = \overline{g_2}$, therefore $f(\lambda w + \mu w') = \lambda f(w) + \mu f(w')$ for any $w, w' \in W$. Hence the lemma holds. □

Since $\widetilde{\mathcal{S}} = \mathrm{cl}_{\widetilde{V}}(\mathcal{S})$ is compact, $\widetilde{\mathcal{S}}$ is also closed in $W$, therefore $\mathrm{cl}_W(\mathcal{S}) = \widetilde{\mathcal{S}}$. Similarly, we have $\mathrm{cl}_{W'}(\mathcal{S}') = \widetilde{\mathcal{S}}'$. Since $f : W \to W'$ is a homeomorphism and $f(\mathcal{S}) = \mathcal{S}'$, we have $f(\widetilde{\mathcal{S}}) = \widetilde{\mathcal{S}}'$. Moreover, since $f : W \to W'$ is an affine isomorphism, $\widetilde{V} = \mathrm{Aff}(\widetilde{\mathcal{S}})$, and $\widetilde{V}' = \mathrm{Aff}(\widetilde{\mathcal{S}}')$, we have $f(\widetilde{V}) = \widetilde{V}'$. Thus $f|_{\widetilde{V}} : \widetilde{V} \to \widetilde{V}'$ is the desired map specified in Theorem 2.1. Hence the proof of Theorem 2.1 is concluded.

## References

[1] Barnum, H., Barrett, J., Leifer, M., Wilce, A.: Generalized no-broadcasting theorem. Phys. Rev. Lett. 99, 240501 (2007). See also arXiv:quant-ph/0611295 (2006)

[2] Barnum, H., Caves, C.M., Fuchs, C.A., Jozsa, R., Schumacher, B.: Noncommuting mixed states cannot be broadcast. Phys. Rev. Lett. 76, 2818 (1996)

[3] Ban, M., Kurokawa, K., Momose, R., Hirota, O.: Optimum measurements for discrimination among symmetric quantum states and parameter estimation. Int. J. Theor. Phys. 36, 1269–1288 (1997)

[4] Bourbaki, N.: Éléments de Mathématique, Topologie Générale 1–4, Second Edition. Masson, Paris (1990)

[5] D'Ariano, G.M.: Probabilistic theories: What is special about quantum mechanics? To appear in: Bokulich, A., Jaeger, G. (eds.) Philosophy of Quantum Information and Entanglement. Cambridge University Press, Cambridge. See also arXiv:0807.4383 (2008)

[6] Dieks, D.: Communication by EPR devices. Phys. Lett. A 92, 271 (1982)

[7] Fuchs, C.A.: Distinguishability and Accessible Information in Quantum Theory. Ph.D. Dissertation, University of New Mexico (1996). See also arXiv:quant-ph/9601020 (1996)

[8] Gudder, S.P.: Quantum Probability. Academic, New York (1988)

[9] Gudder, S.P.: Stochastic Method in Quantum Mechanics. Dover, New York (1979)

[10] Helstrom, C.W.: Quantum Detection and Estimation Theory. Academic, New York (1976)

[11] Holevo, A.S.: Probabilistic and Statistical Aspects of Quantum Theory. Elsevier, Amsterdam (1982)

[12] Ježek, M., Řeháček, J., Fiurášek, J.: Finding optimal strategies for minimum-error quantum-state discrimination. Phys. Rev. A 65, 060301 (2002)

[13] Kimura, G., Miyadera, T., Imai, H.: Optimal state discrimination in general probabilistic theories. Phys. Rev. A 79, 062306 (2009)

[14] Mackey, G.W.: Mathematical Foundations of Quantum Mechanics. Addison-Wesley, Massachusetts (1963)

[15] Nielsen, M.A., Chuang, I.L.: Quantum Computation and Quantum Information. Cambridge University Press, Cambridge (2000)

[16] Ozawa, M.: Optimal measurements for general quantum systems. Rep. on Math. Phys. 18, 11–28 (1980)

[17] Rivest, R.L., Shamir, A., Adleman, L.M.: A method of obtaining digital signatures and public-key cryptosystems. Commun. of the ACM 21, 120–126 (1978)

[18] Schaefer, H.H., Wolff, M.P.: Topological Vector Spaces, Second Edition. Springer-Verlag, Heidelberg (1999)

[19] Shor, P.W.: Algorithms for quantum computation: Discrete logarithms and factoring. SIAM J. on Comput. 26, 1484–1509 (1997)

[20] Wootters, W.K., Zurek, W.H.: A single quantum cannot be cloned. Nature 299, 802–803 (1982)

[21] Yuen, H.P.: Amplification of quantum states and noiseless photon amplifiers. Phys. Lett. A 113, 405–407 (1986)

[22] Yuen, H.P., Kennedy, R.S., Lax, M.: Optimum testing of multiple hypotheses in quantum detection theory. IEEE Trans. Inf. Theory 21, 125–134 (1975)